

STIC-Biotech/ChemLib

7/10/75

From: Davis, Minh-Tam
Sent: Tuesday, July 16, 2002 4:31 PM
To: STIC-Biotech/ChemLib
Subject: Search request for 09/821812

Please search in commercial database and in issued patented files:

1) SEQ ID NO:5

2) Oligomer search for SEQ ID NO:5.

Thank you.

MINH TAM DAVIS

ART UNIT 1642, ROOM 8A01, MB 8E12

305-2008

paty date
3/01

Edward Hart
Technical Info. Specialist
STIC/Biotech
CM 6B02 Tel: 305-8203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 7/17/02
Date Completed: 7/18/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: 2
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 62
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 16:44:24 ; Search time 34.15 Seconds

(without alignments)
1746.607 Million cell updates/sec

Title: US-09-821-812-5

Perfect score: 537

Sequence: 1 GGLTFTVIGTRLGVDPRRLS.....PDHDTQHLIKOLRQHFAML 537

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	187	34.8	260	22	AA041765 Human polypeptide
2	166	30.9	220	21	AA053386 Human colon cancer
3	141	26.3	142	22	ABG18793 Novel human diagno
4	92	17.1	550	22	ABG18795 Novel human diagno
5	64	11.9	207	22	ABG18794 Novel human diagno
6	38	7.1	90	22	AA039979 Human polypeptide
7	16	3.0	101	22	AA032895 Novel human secret
8	8	1.5	123	20	AA093405 Human HEV ORF 3 pr
9	8	1.5	487	21	AA010901 S. xylosus D1ca pr
10	7	1.3	29	17	AA000292 Penicillin V amido
11	7	1.3	36	20	AA097535 Antigenic site of

12	7	1.3	66	21	AA033405 Pinus radiata tran
13	7	1.3	75	22	AA034866 E. coli cellular p
14	7	1.3	75	22	AA035457 Haemophilus influe
15	7	1.3	75	22	AA038266 Salmonella typhi c
16	7	1.3	75	22	AA038266 Human nervous syst
17	7	1.3	82	22	AA042302 Propionibacterium
18	7	1.3	103	20	AA019803 B. burgdorferi ant
19	7	1.3	133	20	AA019802 B. burgdorferi ant
20	7	1.3	144	21	AA026620 Arabidopsis thalia
21	7	1.3	146	21	AA048630 Arabidopsis thalia
22	7	1.3	146	21	AA033260 Pinus radiata tran
23	7	1.3	171	22	AB060190 Drosophila melanog
24	7	1.3	209	22	AB071753 Arabidopsis thalia
25	7	1.3	213	21	AA026619 Arabidopsis thalia
26	7	1.3	213	21	AA048629 Arabidopsis thalia
27	7	1.3	215	15	AA060575 House dust mite al
28	7	1.3	215	20	AA025586 D. pteronyssinus a
29	7	1.3	267	19	AA081726 M. tuberculosis im
30	7	1.3	267	19	AA064359 Mycobacterium tube
31	7	1.3	267	20	AA039156 M. tuberculosis an
32	7	1.3	267	20	AA039013 M. tuberculosis re
33	7	1.3	287	22	AB073558 Novel human diagno
34	7	1.3	290	21	AA023634 Arabidopsis thalia
35	7	1.3	307	21	AA023633 Arabidopsis thalia
36	7	1.3	309	21	AA074359 Neisseria gonorrhe
37	7	1.3	309	21	AA074361 Neisseria meningit
38	7	1.3	310	21	AA023632 Arabidopsis thalia
39	7	1.3	312	21	AA074360 Neisseria meningit
40	7	1.3	315	21	AA058367 Anabaena variabil
41	7	1.3	321	22	AB020636 Novel human diagno
42	7	1.3	340	21	AA052000 Human AC11 protein
43	7	1.3	340	21	AA051629 Putative P. abyssal
44	7	1.3	363	22	AA086616 Eucalyptus grandis
45	7	1.3	370	21	AA032774

ALIGNMENTS

RESULT 1	
ID	AA041765 standard; Protein: 260 AA.
XX	
AC	AA041765:
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 6696.
XX	
KW	Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	anyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia.
XX	
OS	Homo sapiens.
XX	
PN	WO20015312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	

PA (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR N-PSDB; AAI60921.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS Example 2: SEQ ID NO 6696; 10078pp; English.
 XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic.
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemoclastic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX Sequence 260 AA;
 SQ
 Query Match 34.8%; Score 187; DB 22; Length 260;
 Best Local Similarity 100.0%; Pred. No. 5,4e-183;
 Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 30 PAAAMSEMEPLLAWSYFRRRKPOLCADLCTQMLEKSPYDQAAWILKARALTEMVYIDEI 89
 DB 17 PAAAMSEMEPLLAWSYFRRRKPOLCADLCTQMLEKSPYDQAAWILKARALTEMVYIDEI 76
 QY 90 DVDEGIAEWMLEDAIAQVPRPGTSIKLPETNOTGSPSOAVRPIITQAGRTIFLRPST 149
 DB 77 dvdegiagmldenaiaqvprrgtsiklpqtnqigpssqavrpilqagrpilqfipst 136
 QY 150 QSGRPGTMEQAIRPRAYARPRPTSSGFRVRLGTASMLTSPGPFINTSLRLULTYSSQ 209
 DB 137 qsgrrpgtmeqairprayarprptssgfrvrlgtasmltspgpfintslrlultkysq 196
 QY 210 KPRLAKA 216
 DB 197 kpriaka 203
 RESULT 2
 AAB53386
 ID AAB53386 standard; Protein: 220 AA.
 XX
 XX AAB53386;
 DT 09-MAR-2001 (first entry)
 DE Human colon cancer antigen protein sequence SEQ ID NO:926.
 XX
 XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytosolic; cardiovascular; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.
 XX

OS Homo sapiens.
 XX
 XX W0200055351-A1.
 XX
 PD 21-SEP-2000.
 XX
 XX 08-MAR-2000; 2000WO-0505883.
 PF
 XX 12-MAR-1999; 99US-0124270.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI WPI: 2000-587534/55.
 DR N-PSDB; AAC98143.
 XX
 PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 XX
 XX Claim 11: Page 1485; 2104pp; English.
 PS
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytosolic, cardiovascular, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders, infectious
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 XX
 XX Sequence 220 AA;
 SQ
 Query Match 30.9%; Score 166; DB 21; Length 220;
 Best Local Similarity 100.0%; Pred. No. 1,6e-161;
 Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 372 RRLQMGVINGQLFNINIGLCCEFYAQQYDMLTSPERALSIAENESEAADVYNIQHVAVG 431
 DB 55 rrlqmglyngqlfninlgclcfyagqymltstferalsiaeneeeadvynlghvav 114
 QY 432 IGDFTNLHQCFRLALVNNNNHAEAYNNLAVLEMKRGHYEQARALQTAASSLAPHMYEPHF 491
 DB 115 lgdftnlhqcfrrlalnvnnnhaeaynnlavllemkrgyveqarallqtaasslaphmyepf 174
 QY 492 NFATISDKIGDLQRSYVAQAQSEAAFPDHDVTOHLIKOLRQHPML 537
 DB 175 nfatisdkigdlqrsyvaqaqseaafpdhvdtghliklqrghfaml 220
 RESULT 3
 ABG18793
 ID ABG18793 standard; Protein: 142 AA.
 XX
 XX ABG18793;
 AC
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #18784.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 KW
 XX

OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB: AAS82980.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 49152; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 142 AA;
 XX

Query Match 26.3%; Score 141; DB 22; Length 142;
 Best Local Similarity 100.0%; Pred. No. 4.5e-136;
 Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 YFRRRFOICADLCOTOMLEKSPYQAMLIKARALTEMVYIDETIDVQESIAEMLDENA 105
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 2 yfrrrfqicadlctgmlekydgaawlikaraltemyideldvdgqesiaemldena 61
 QY 106 IAOVPPRGSLKLGSTNOCGPGSAOAVRPTQACRPITGRLRPSSTOSGRPTMGQAIRTPR 165
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 62 laqvprgtslklpgtngtqgsqavrpilgtlgrpistqsgsrptmeqairtptr 121
 QY 166 TATARTPTSSSGRFRIGTA 186
 ||||||||||||||||||||||||
 DB 122 tatyartptsssgrrivrigna 142

RESULT 4
 ABG18795
 ID ABG18795 standard; Protein; 550 AA.
 XX
 AC ABG18795;
 XX

DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #18786.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB: AAS82982.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 49154; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 550 AA;
 XX

Query Match 17.1%; Score 92; DB 22; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2.8e-85;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 AIAICGSNHFYSDOPETALRFYRLQMGTYNGOLFNNLGLCCPYAODYMTLTSFERAL 409
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 361 aiaicgsnhfysdopetalfyrrllimglyngqlfnnlglcclyaqdydmlltsferal 420
 QY 410 SLAENEEDADWYNLGHVAVGIDTNLAHQ 441
 ||||||||||||||||||||||||
 DB 421 slaeneeadwvynlghvavgidtnlahqc 452

RESULT 5
 ABG18794
 ID ABG18794 standard; Protein; 207 AA.
 XX

XX AEG18794;
AC
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #18785.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
PD
XX 11-OCT-2001.
PF 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Dermanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PDB; AAS82981.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS
XX Claim 20; SEQ ID NO 49153; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
CC
XX
SQ Sequence 207 AA;

Query Match 11.9%; Score 64; DB 22; Length 207;
Best Local Similarity 100.0%; Pred. No. 5.7e-57;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 QAAMWIKARALTEMVYDEIDVOEGIAEMMLDENATAOVRPRTSLKLGCTNOTGSPSQ 129
DB 1 qaawwikaraltemvydeidvoegiaemmldenataovrprrtsliklgctnotgspsq 60
OY 130 AVRP 133
DB 61 avrp 64

RESULT 6
AAM3979
ID AAM3979 standard; Protein: 90 AA.
XX
XX AAM3979;
AC
XX
XX 22-OCT-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 3124.
DE
XX
XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
OS
XX
XX WO200153312-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000MO-US34263.
PF
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
XX 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
XX
XX 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
XX
XX 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dermanac RT;
XX
XX WPI; 2001-442253/47.
DR N-PDB; AAI59135.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PT
XX
XX Example 4; SEQ ID NO 3124; 10078bp; English.
PS
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA42213) with noctropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/Inhibin activity, chemotactic/chemokine activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
CC
XX
SQ Sequence 90 AA;

Query Match 7.1%; Score 38; DB 22; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.2e-30;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 377 MGIYNGOLFNNLGLCFYAAOYDMTLTSFBRALSLAEN 414

Db 51 mglyngqlfmlgclccfyagqymtltsferalslaen 88

RESULT 7

AAU32896
ID AAU32896 standard; Protein; 101 AA.
XX
AC AAU32896;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #3387.
XX
KW Human: vaccination; gene therapy; nutritional supplement;
KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN MO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001MO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
XX
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR MPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy -
XX
PS Claim 20; Page 685; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 101 AA;

Query Match 3.0%; Score 16; DB 22; Length 101;
Best Local Similarity 100.0%; Pred. No. 4.9e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 EAYNNLAVLEMRKGV 469
Db 30 eaynnlavlemrkghv 45

RESULT 8
AAW93405
ID AAW93405 standard; Protein; 123 AA.

XX
AC AAW93405;

XX 11-JUN-1999 (first entry)

XX Human HEV ORF 3 protein from strain Helian.

XX Swine hepatitis E virus; HEV; cross-reaction; antibody; human; therapy;
KM vaccine; immunise; infection; detection; diagnosis; prevention.

XX Hepatitis E virus.

XX WO9904029-A2.

XX 28-JAN-1999.

XX 17-JUL-1998; 98WO-US14665.

XX 18-JUL-1997; 97US-0053069.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Emerson SU, Meng X, Purcell RH;

XX MPI; 1999-132270/11.

XX New isolated swine hepatitis E virus - used to develop products for
PT the diagnosis, prevention and treatment of hepatitis E virus
PT infection in mammals, particularly humans

XX Example 1; Fig 3B; 70pp; English.

CC This invention describes a swine hepatitis E virus (HEV) and its natural
CC mutants which are capable of cross-reacting with antibodies reactive
CC with a human HEV strain or natural mutants. The HEV and the proteins
CC can be used in vaccines for immunising against HEV infection. The swine
CC HEV can be used in humans to prevent possible infection by human HEV. The
CC swine HEV can also be used as a therapeutic treatment for infection by
CC other strains of HEV. The swine HEV can also be used for the production
CC of antibodies which can be used in therapy, detection and diagnosis. The
CC products can also be used for determining the susceptibility of cells or
CC organs to infection with swine HEV. The swine HEV is particularly useful
CC for the development of agents for the prevention, treatment and detection
CC of human HEV because it is not a human virus and thus can be handled both
CC experimentally and clinically without fear of severe infection and/or
CC contamination.

XX Sequence 123 AA;

Query Match 1.5%; Score 8; DB 20; Length 123;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 MNMSSAA 334
Db 1 mnmssaa 8

RESULT 9

AAAB10901
ID AAB10901 standard; Protein; 487 AA.

XX
AC AAB10901;

XX 30-JAN-2001 (first entry)

XX S. xylosus DltA protein.

XX dltA; dltB; dltC; dltD; dltABCD operon; antibacterial; D-alanine;
KM telcholic acid; endotoxin-like; inflammation; gram-positive bacteria;
XX antimicrobial; D-alanine-D-alanyl carrier protein ligase; ds.

OS Staphylococcus xylosus.
 XX DE19912706-A1.
 XX 07-SEP-2000.
 PD 20-MAR-1999; 99DE-1012706.
 XX 05-MAR-1999; 99DE-1009636.
 XX (PSTR-) PETRY GENMEDICS GMBH.
 PA Goetz F, Peschel A;
 PI WPI; 2000-588432/56.
 DR New staphylococcal DNA for dltABCD operons, useful e.g. for identifying
 PT antibacterials and agents that reduce bacterial resistance to
 PT antimicrobials -
 XX Claim 4; Fig 5; 20pp; German.
 PS
 XX This invention describes novel DNA sequences (A) of the dltABCD operon
 CC from Staphylococcus xylosus and Staphylococcus aureus. The products of
 CC the invention have antibacterial activity. The proteins expressed by (A)
 CC are involved in incorporation of D-alanine (Dala) into teichoic acid (I)
 CC (which has endotoxin-like inflammatory activity) by Gram-positive
 CC bacteria. Incorporation of Dala into (I) is correlated with sensitivity
 CC of bacteria to antimicrobial agents. I.e. Dala is necessary for
 CC resistance. When tested against wild-type *S. aureus* Sall3, the human
 CC neutrophilic peptide defensin had minimum inhibitory concentration (MIC)
 CC over 100 micro g/ml, but against a mutant in which the dltA gene has been
 CC deleted it had MIC over 10 micro g/ml. Similar reductions in MIC were
 CC determined for other cationic antibacterial peptides. (A), optionally
 CC mutated, are used to study the function of their encoded proteins.
 CC Involved in incorporation of D-alanine (Dala) into teichoic acid (I).
 CC Agents that reduce the inflammatory activity of (I) or incorporation of
 CC Dala into (I) are used: (i) to increase the sensitivity of Gram-positive
 CC bacteria to antimicrobial agents; (ii) to inhibit formation of biofilms
 CC (particularly of staphylococci) on glass, metal or plastics surfaces
 CC (e.g. catheters or cardiac pacemakers); and (iii) as antibacterials. This
 CC sequence represents the Staphylococcus xylosus D-alanine-D-alanyl carrier
 CC protein ligase (dltA) which is described in the method of the invention.
 CC
 XX
 SQ Sequence 487 AA:
 Query Match 1.5%; Score 8; DB 21; Length 487;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 409 LSLAENEK 416
 |||||
 Db 322 LSLAENE 329
 RESULT 10
 AAW00292
 ID AAW00292 standard; Peptide; 29 AA.
 XX
 AC AAW00292;
 XX
 DT 20-NOV-1996 (first entry)
 XX
 DE Penicillin V amidohydrolyase N-terminus.
 XX
 KM Penicillin V amidohydrolyase; PVA; F. oxysporum; strain 435;
 KM hydrolysis: phenoxymethylpenicillin; 6-aminopenicillanic acid;
 KM 6-APA; beta-lactam; semi-synthetic penicillin; expression vector;
 KM recombinant production.
 XX
 OS Fusarium oxysporum.
 XX

FN Key Location/Qualifiers
 FT Misc-difference 7
 FT /Label- Ala, Lys
 FT Misc-difference 9
 FT /note- "Any amino acid"
 FT Misc-difference 22
 FT /Label- Thr, Val
 XX
 PN US516679-A.
 XX 14-MAY-1996.
 PD
 XX
 XX 23-DEC-1994; 94US-0363475.
 XX
 XX 23-DEC-1994; 94US-0363475.
 PR
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA
 XX Burnett WV, Chiang S, Tonzi SM;
 PI WPI; 1996-251011/25.
 DR
 XX
 XX DNA encoding penicillin V amidohydrolyase (PVA) from Fusarium
 PT oxysporum - also recombinant vectors and host cells for production
 PT of PVA for use in the manufacture of penicillin.
 PS
 XX Claim 18; Fig 1; 46pp; English.
 CC The sequences given in AAW00292-98 and AAW00300 are peptide fragments
 CC derived from the secreted form of penicillin V amidohydrolyase (PVA)
 CC from F. oxysporum strain 435. The secreted form of PVA is a
 CC glycoprotein of mol. wt. 65 kD. The seven amino acid fragment of
 CC peptide C given in AAW00298 was used in the design of four probes by
 CC reverse translation (see also AAT40247). This probe was used in the
 CC identification of a PVA cDNA clone. PVA is used for the enzymatic
 CC hydrolysis of penicillin V (phenoxymethylpenicillin) to 6-amino-
 CC penicillanic acid (6-APA). 6-APA is the active beta-lactam nucleus
 CC used in the manufacture of semi-synthetic penicillins. The PVA coding
 CC sequences can be inserted into expression vectors for the recombinant
 CC production of PVA in a suitable host, pref. Fusarium sp.
 CC
 XX
 SQ Sequence 29 AA:
 Query Match 1.3%; Score 7; DB 17; Length 29;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 115 SLKRPET 121
 |||||
 Db 15 SLKIPGT 21
 RESULT 11
 AAW97535
 ID AAW97535 standard; peptide; 36 AA.
 XX
 AC AAW97535;
 XX
 DT 19-MAY-1999 (first entry)
 XX
 DE Antigenic site of HN protein loop beta-5L01.
 XX
 KM Antigenic site; haemagglutinin-neuraminidase; HN; paramyxoviridae;
 KM virus epitope; attachment protein; vaccine; immunodominant epitope.
 XX
 OS Simian virus 9.
 OS
 XX WO9902695-A2.
 PN
 XX 21-JAN-1999.
 PD
 XX 08-JUL-1998; 98WO-NL00390.
 PF

CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 S0 Sequence 75 AA;

Query Match 1.38; Score 7; DB 22; Length 75;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 46 YFRRRRP 52
 |||||
 Db 4 yfrrrkf 10

Search completed: July 17, 2002, 16:46:48
 Job time: 144 sec

THIS PAGE BLANK (USPTO)


```

Db 232 EEAAPVWNLGVAVAGIDDTNLAHOCFRLALVNNNNHAEVYNNNLVLEMRKGVHQAARL 291
Oy 476 IOTASSLAPHHYEPHFENFATISDKIGDLORSVVAQAOKSEAPDPDHVQHIKOLROHA 535
Db 292 IOTASSLAPHHYEPHFENFATISDKIGDLORSVVAQAOKSEAPDPDHVQHIKOLROHA 351
Oy 536 ML 537
Db 352 ML 353

RESULT 2
OyDCP7 PRELIMINARY; PRT; 505 AA.
ID 09DCP7;
AC 09DCP7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 0610012F22RIK PROTEIN.
GN 0610012F22RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shuegawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arkawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
Kushl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
Schrini L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boiffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustinch S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,
Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
EMBL: AK002597; BAB22218.1;
DR MGI: MGI:1923510; 0610012F22RIK.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 5.
DR SMART: SMO0028; TPR; 7.
SO SEQUENCE 505 AA; 57405 MW; 9B13D8F9D9E4F22 CRC64;

Query Match 16.4%; Score 88; DB 11; Length 505;
Best Local Similarity 100.0%; Pred. No. 4,9e-82;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 129 QAVRPITGAGRPITGFLRPSTQSGRPQTMEQAIRPTAYTARPTITSSSGRFRVLGTASM 188
Db 97 QAVRPITGAGRPITGFLRPSTQSGRPQTMEQAIRPTAYTARPTITSSSGRFRVLGTASM 156
Oy 189 LTPSPGPFINLSRLNLTKYSQKPKLAKA 216
Db 157 LTPSPGPFINLSRLNLTKYSQKPKLAKA 184

RESULT 3
OyDCP7 PRELIMINARY; PRT; 123 AA.
ID 081870
Oy 081870

```

```

AC 081870;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ORF 3 PRECURSOR.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
CX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
Fry K.E., Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
full-length viral genome.";
RL Virology 185:120-131(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92261377; PubMed=1584074;
RA Uchida T., Suzuki K., Hayashi N., Iida F., Hara T., Oo S.S.,
Wang C.-K., Shikata T., Ichikawa M., Rikihisa T., Mizuno K., Win K.M.;
RT "Hepatitis E virus: cDNA cloning and expression.";
RL Microbiol. Immunol. 36:67-79(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92335008; PubMed=1630924;
RA Aye T.T., Uchida T., Ma X.Z., Iida F., Shikata T., Zhuang H.,
Win K.M.;
RT "Complete nucleotide sequence of a hepatitis E virus isolated from the
Xinjiang epidemic (1986-1988) of China.";
RL Nucleic Acids Res. 20:3512-3512(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=92115700; PubMed=1731327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
Malik I.A., Iqbal M., Purcell R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92271462; PubMed=1589964;
RA Fry K.E., Tam A.W., Smith M.W., Kim J.P., Luk K.-C., Young L.M.,
Pitak M., Feldman R.A., Yun K.Y., Purdy M.A., McCaustland K.A.,
Bradley D.W., Reyes G.R.;
RT "Hepatitis E virus (HEV): Strain variation in the nonstructural gene
region encoding consensus motifs for an RNA-dependent RNA polymerase
and an ATP/GTP binding site.";
RL Virus Genes 6:173-185(1992).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=93348763; PubMed=8346669;
RA Bi S.L., Purdy M.A., McCaustland K.A., Margolis H.S., Bradley D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China.";
RL Virus Res. 28:233-247(1993).
EMBL: L08816; AAA03190.1;
DR InterPro: IPR003384; HEV_ORF2.
DR Pfam: PF02444; HEV_ORF2; 1.
KW Signal.
FT SIGNAL 14 32 POTENTIAL.
SO SEQUENCE 123 AA; 12588 MW; 92F59ED6B49C1755 CRC64;

Query Match 1.5%; Score 8; DB 12; Length 123;
Best Local Similarity 100.0%; Pred. No. 8,2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 327 MNNSSSAA 334
Db 1 MNNSSSAA 8

RESULT 4

```

092MG2
ID 092MG2 PRELIMINARY; PRT; 337 AA.
AC 092MG2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEICAL TRANSMEMBRANE PROTEIN SMC00712.
GN SMC00712.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1021;
RX MEDLINE-21368234; PubMed-11474104;
RA Galibert F., Finan T.M., Long S.R., Pehler A., Abola P., Ampe F.,
Barloy-Hubier F., Barnett M.J., Becker A., Bolstad P., Bothe G.,
Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
Hernandez-Lucias I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
Ramsberger U., Surzycki R., Thebaud P., Vandendol M.,
Vornheller F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RA "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RT Science 293:668-672(2001).
RL EMBL; AL591791; CAC47239.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 337 AA; 35599 MW; 99CPCF915F896BF CRC64;

Query Match 1.5%; Score 8; DB 16; Length 337;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 24 GPSLAPA 31
Db 201 GPSLAPA 208

RESULT 5
09X2N4
ID 09X2N4 PRELIMINARY; PRT; 487 AA.
AC 09X2N4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE D-ALANINE-D-ALANYL CARRIER PROTEIN LIGASE.
GN DLA.
OS Staphylococcus xylosus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1288;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C2A;
RX MEDLINE-99185055; PubMed-10085071;
RA Peschel A., Otto M., Jack R.W., Kalbacher H., Jung G., Gotz F.;
RT "Inactivation of the dlt operon in Staphylococcus aureus confers
RT sensitivity to defensins, protegrins, and other antimicrobial
RT peptides.";
RT J. Biol. Chem. 274:8405-8410(1999).
RL EMBL; AF032440; AAD01942.1; -
DR HSP; P14687; IAMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR000977; DNA_ligase.
DR Pfam; PF00501; AMP-binding; 1.
DR PRINT; PR00154; AMPBINDING.
DR PROSITE; PS00697; DNA_LIGASE_A1; UNKNOWN_1.
KW ligase.
SQ SEQUENCE 487 AA; 55729 MW; 283C975BC59E8BD2 CRC64;

Query Match 1.5%; Score 8; DB 2; Length 487;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 409 LSLAENE 416
Db 322 LSLAENE 329

RESULT 6
095XK1
ID 095XK1 PRELIMINARY; PRT; 772 AA.
AC 095XK1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEICAL 87.6 KDA PROTEIN.
GN Y102A11A.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RT Science 282:2012-2018(1998).
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Lamar B., Fulton B., Minx P., Haakenson W., Elliott G., Gregory S.;
RT "The sequence of C. elegans cosmid Y102A11A.";
RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084152; AAK39314.2; -
KW Hypothetical protein.
SQ SEQUENCE 772 AA; 87562 MW; 86EBE1C830C7FDIC CRC64;

Query Match 1.5%; Score 8; DB 5; Length 772;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 522 DTQHLIKQ 529
Db 725 DTQHLIKQ 732

RESULT 7
091ZD2
ID 091ZD2 PRELIMINARY; PRT; 1517 AA.
AC 091ZD2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CCAAT DISPLACEMENT PROTEIN CDP.
GN CUT11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE-21429086; PubMed-11544187;
 RA Ellis T., Gambardella L., Horcher M., Tschanz S., Capol J.,
 RA Bertram P., Jochum W., Barrandon Y., Busslinger M.;
 RT "The transcriptional repressor CDP (Ctll1) is essential for epithelial
 RT cell differentiation of the lung and the hair follicle."
 RL Genes Dev. 15:2307-2319(2001).
 DR EMBL:AY037807; AAK59986.1; -
 SQ SEQUENCE 1517 AA; 165852 MW; B17A8740621EBBAC CRC64;

Query Match 1.5%; Score 8; DB 11; Length 1517;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 PSLAAPAA 32
 DB 1415 PSLAAPAA 1422

RESULT 8
 ID 093TW9 PRELIMINARY; PRT; 5192 AA.
 AC 093TW9;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE MAC.
 GN MAC.
 OS Stigmatella aurantiaca.
 OC Bacteria; Proteobacteria; delta subdivision: Myxobacteria;
 CC Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
 OX NCBI_TaxID=41;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21110452; PubMed-11182319;
 RA Sliakowski B., Nordsiek G., Kunze B., Blocker H., Muller R.;
 RT "Novel features in a combined polyketide synthase/non-ribosomal
 RT peptide synthetase: the myxalimid biosynthetic gene cluster of the
 RT myxobacterium Stigmatella aurantiaca Sga15."
 RL Chem. Biol. 8:59-69(2001).
 DR EMBL:AF319998; AAK57187.1; -
 SQ SEQUENCE 5192 AA; 558274 MW; 7C7CAAE414A31DAE CRC64;

Query Match 1.5%; Score 8; DB 2; Length 5192;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 PFERALSL 411
 DB 1335 PFERALSL 1342

RESULT 9
 ID 083097 PRELIMINARY; PRT; 71 AA.
 AC 083097;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE RNA-DEPENDENT RNA POLYMERASE (FRAGMENT).
 GN RRP.
 OS Leishmania RNA virus 1-13.
 OC Viruses; dsRNA viruses; Totiviridae; Leishmanaviruses.
 OX NCBI_TaxID=39117;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95334386; PubMed-7610059;
 RA Wildmer G., Dooley S.;
 RT "Phylogenetic analysis of Leishmania RNA virus and Leishmania suggests
 RT ancient virus-parasite association."
 RL Nucleic Acids Res. 23:2300-2304(1995).
 DR EMBL: L39069; AAC42114.1; -

KW RNA-directed RNA polymerase.
 FT NON_TER 1
 FT NON_TER 71
 SQ SEQUENCE 71 AA; 8266 MW; F3A8FB7EAB8B563F CRC64;

Query Match 1.3%; Score 7; DB 12; Length 71;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 88 EIDVDOE 94
 DB 17 EIDVDOE 23

RESULT 10
 ID 098IF7 PRELIMINARY; PRT; 81 AA.
 AC 098IF7;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DE MSR2423 PROTEIN.
 DE MSR2423.
 GN MSR2423.
 OS Rhizobium loti (Mesorhizobium loti)
 OC Bacteria; Proteobacteria; alpha subdivision: Rhizobiaceae group;
 CC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=MAFF303099;
 RX MEDLINE-21082930; PubMed-11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Tanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti."
 RT Mesorhizobium loti.
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP002999; BAB49559.1; -
 DR InterPro: IPR002145; CofG_HTH_4.
 DR Pfam: PF01402; HTH_4; 1.
 KW Complete proteome.
 SQ SEQUENCE 81 AA; 9035 MW; 7652A22945D0BFC7 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 81;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 505 RSYVAAQ 511
 DB 29 RSYVAAQ 35

RESULT 11
 ID 09EAZ3 PRELIMINARY; PRT; 87 AA.
 AC 09EAZ3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE GP120 (FRAGMENT).
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A324.
 RA Machuca R.A.R., Bogh M., Gersloft J., Kvinesdal B., Pedersen C.,
 RA Odel N., Nielsen H., Nielsen C.;

RT "HIV-2 subtypes in Denmark."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A324;
 RA Bogn M., Machuca R.A.R., Nielsen C.;
 RT "Subtype specific problems with Roche PCR."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ89467; CAC06444.1; -
 DR InterPro: IPR000777; GP120.
 DR Pfam: PF00516; GP120; 1.
 KW AIDS; Coat protein; Glycoprotein.
 FT NON_TER 1
 FT NON_TER 87
 SQ SEQUENCE 87 AA; 9917 MW; 11415473F6861035 CRC64;

Query Match 1.3%; Score 7; DB 15; Length 87;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 528 KOLROHF 534
 DB 76 KOLROHF 82

RESULT 12
 ID 090T11 PRELIMINARY; PRT; 97 AA.
 AC 090T11;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ORF 5
 OS SVTS2, plectrovirus.
 OC Viruses; ssDNA viruses; Inoviridae; Plectrovirus.
 NCBI_TaxID=93224;
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20227373; PubMed-10766306;
 RA Sna Y., Melcher U., Davis R.E., Fletcher J.;
 RT "Common elements of spiroplasma plectroviruses revealed by nucleotide
 RT sequence of SVTS2.";
 RL Virus Genes 20:47-56(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Melcher U.K., Sna Y., Davis R.E., Fletcher J.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF133242; AAF18313.1; -
 SQ SEQUENCE 97 AA; 10656 MW; 5E1C77293FE2870B CRC64;

Query Match 1.3%; Score 7; DB 12; Length 97;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 LKLRGTN 122
 DB 31 LKLRGTN 37

RESULT 13
 ID 093RB1 PRELIMINARY; PRT; 110 AA.
 AC 093RB1;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE FARNESYL DIPHOSPHATE SYNTHASE (FRAGMENT).
 GN PPPS.
 OS Rhodovulum strictum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodovulum.

OX NCBI_TaxID=56314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 9220;
 RA Cantlera J.L., Kawasaki H., Seki T.;
 RT "Molecular systematic studies of phototrophic bacteria using farnesyl
 RT diphosphate synthase.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB053177; BAB61883.1; -
 FT NON_TER 1
 FT NON_TER 110
 SQ SEQUENCE 110 AA; 11669 MW; A715AB3622BABC27 CRC64;

Query Match 1.3%; Score 7; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 407 RALSIAE 413
 DB 86 RALSIAE 92

RESULT 14
 ID 09RV55 PRELIMINARY; PRT; 116 AA.
 AC 09RV55;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHEETICAL 12.3 KDA PROTEIN.
 GN DR0947.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-R1;
 RX MEDLINE-20036896; PubMed-10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fieischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001947; AAF10529.1; -
 DR TIGR; DR0947; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 116 AA; 12298 MW; 168391024730A71E CRC64;

Query Match 1.3%; Score 7; DB 16; Length 116;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 468 HVEQARA 474
 DB 8 HVEQARA 14

RESULT 15
 ID 09YEB1 PRELIMINARY; PRT; 116 AA.
 AC 09YEB1;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHEETICAL 12.6 KDA PROTEIN APE0663.
 GN APE0663.

OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1.
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankal A., Kosugi H.,
 Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
 Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000060; BAA79635.1; -
 KW Hypothetical protein; complete proteome.
 SO SEQUENCE 116 AA; 12585 MW; 237AC678B52A85CF CRC64;

Query Match

1.3%; Score 7; DB 17; Length 116;

Best Local Similarity 100.0%; Pred. No. 84;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 177 SGRFVRL 183
|||||||

DB 20 SGRFVRL 26

Search completed: July 17, 2002, 16:48:16
 Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 16:43:24 ; Search time 30.75 Seconds

(without alignments)
3021.082 Million cell updates/sec

Title: US-09-821-812-5

Perfect score: 2797

Sequence: 1 GGLTFTVIGTRLGVDPRRLS.....PDHVDTOHLIKQLRQHFAML 537

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2443	87.3	505	11	Q9DCP7 mus musculi
2	1733	62.0	353	4	Q96DG8 homo sapien
3	950.5	34.0	458	5	Q23049 caenorhabdi
4	576.5	20.6	549	5	Q9VPP9 drosophila
5	194	6.9	379	17	Q26186 methanot
6	194	6.9	1011	5	Q9Y148 drosophila
7	194	6.9	1059	5	Q9V3X6 drosophila
8	186	6.6	403	17	Q26176 methanot
9	185.5	6.6	464	4	Q96H45 homo sapien
10	185.5	6.6	1036	4	Q96CC1 homo sapien
11	185.5	6.6	1046	11	Q91Y38 mus musculi
12	185.5	6.6	1102	3	Q60184 schizosac
13	183	6.5	977	10	Q9M8Y0 arabidopsi
14	182.5	6.5	519	4	Q96RK4 homo sapien
15	178	6.4	1151	5	Q18158 caenorhabdi
16	168	6.0	782	10	Q9FPW0 chlamydom

17	167	6.0	1173	11	Q62018 mus musculi
18	165.5	5.9	305	16	Q97DM4 clostridium
19	165.5	5.9	1564	4	Q15077 homo sapien
20	164.5	5.9	932	10	Q82039 petunia hyb
21	164	5.9	524	6	Q9BG26 macaca fasc
22	163.5	5.8	1173	4	Q15015 homo sapien
23	160.5	5.7	774	5	Q16296 caenorhabdi
24	159	5.7	1292	3	Q96MLO usllago ma
25	158.5	5.7	1115	10	Q9SIS5 arabidopsi
26	154.5	5.5	238	11	Q9CS83 mus musculi
27	154.5	5.5	479	4	Q96S08 homo sapien
28	151.5	5.4	914	10	Q96301 arabidopsi
29	148	5.3	206	16	Q34452 bacillus su
30	147	5.3	802	10	Q80882 arabidopsi
31	145.5	5.2	944	10	Q82422 hordeum vul
32	145.5	5.2	1390	5	Q77033 dictyostell
33	143	5.1	486	5	Q9V5T0 drosophila
34	143	5.1	948	16	Q55489 synecocyst
35	142.5	5.1	1150	5	Q9M0H4 drosophila
36	142	5.1	429	16	P73344 synecocyst
37	142	5.1	808	10	Q23052 arabidopsi
38	141.5	5.1	370	16	Q9K584 bacillus ha
39	141	5.0	830	4	Q96F35 homo sapien
40	140.5	5.0	339	16	Q927L1 chlamydia p
41	140.5	5.0	645	4	Q96N52 homo sapien
42	140	5.0	1432	3	Q06585 saccharomyc
43	138	4.9	335	16	Q9PLP4 chlamydia m
44	137.5	4.9	280	16	Q96G13 rhizobium l
45	137	4.9	697	10	Q9L016 arabidopsi

ALIGNMENTS

RESULT	ID	Q9DCP7	PRELIMINARY:	PRT:	505 AA.
AC	Q9DCP7				
DT	01-JUN-2001	(TRENBLER, 17, Created)			
DT	01-JUN-2001	(TRENBLER, 17, Last sequence update)			
DT	01-JUN-2001	(TRENBLER, 17, Last annotation update)			
DE	0610012F22R1K	PROTEIN.			
GN	0610012F22R1K				
OS	Mus musculus (Mouse)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
FM	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=KIDNEY;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,				
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,				
RA	Kodota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schirml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Guinacich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,				
RA	Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohatsu S.,				
RA	Hayashizaki Y.,				
RT	"Functional annotation of a full-length mouse cDNA collection.";				
RL	Nature 409:685-690(2001).				
DR	EMBL; AK002597; BAB22218.1; -				
DR	MGD; MGI:1923510; 0610012F22R1K.				

DR InterPro: IPR001440; TPR.
 DR Pfam: PF00515; TPR; 5.
 DR SMART: SM00028; TPR; 7.
 SQ SEQUENCE 505 AA; 57405 MW; 9B13D8FD9DEB4F22 CRC64;

Query Match 87.3%; Score 2443; DB 11; Length 505;
 Best Local Similarity 92.2%; Pred. No. 9,9e-205;
 Matches 471; Conservative 15; Mismatches 13; Indels 12; Gaps 3;

33 MSSEPEPLLAASYFRRRKPOLCADLCTOMLEKSPYDAAAILKARATEWYIDEIDVD 92
 1 MGSEPEPLRAMSYFRRRKPOLCADLCTOMLEKSPYDAAAILKARATEWYIDEIDVD 60
 QY QEGIAEMMLDENALAOVPRPGTSKLPGTNOTGSPQAVRPITQAGREITGFLRSTQSG 152
 61 QEGIAEMMLDENALAOVPRPGTSKLPGTNOTGSPQAVRPITQAGREITGFLRSTQSG 120
 QY 153 RPTMEQAIKRPRTAYTARPTSSSGRFVRLGTASMLTSPDGPRLNLSRLNLTYSQPK 212
 121 RPTMEQAIKRPRTAYTARPTSSSGRFVRLGTASMLTSPDGPRLNLSRLNLTYSQPK 180
 QY 213 LAKACLSISFIM-----KMLRLIMLALSTESQYKDDMMKVQIGKCYRLGMVREA 266
 181 LAKAL--FEYLLHENDVKMALDL---ASLSTESQYKDDMMKVQIGKCYRLGMVREA 234
 QY EKQFSALQOEVDYFLYLAQVYSLDPYALNLFKQGLKPGEVTLTGCIARIYEE 326
 235 EKQFSALQOEVDYFLYLAQVYSLDPYALNLFKQGLKPGEVTLTGCIARIYEE 294
 QY 327 MNNSSAEYKVKELKODNTVHXAIACTGSHFYSDDPEIARFVRLLOGLYNGQLFN 386
 295 MNNSSAEYKVKELKODNTVHXAIACTGSHFYSDDPEIARFVRLLOGLYNGQLFN 354
 QY 387 NLGICCFPAOQYDMTITSFERALSLAESEEAADWYNLGHVAVIGTNTLAHOCFRAL 446
 355 NLGICCFPAOQYDMTITSFERALSLAESEEAADWYNLGHVAVIGTNTLAHOCFRAL 414
 QY 447 VNNNNHAEVYNNLAVLEMRKGVQEARALLQASSLAPHMYEPHFNFATISDKIGDLORS 506
 415 VNNNNHAEVYNNLAVLEMRKGVQEARALLQASSLAPHMYEPHFNFATISDKIGDLORS 474
 QY 507 YVAAKSEAFPPDHVDTQHLIKOLRQHFAML 537
 475 YVAAKSEAFPPDHVDTQHLIKOLRQHFAML 505
 DB
 RESULT 2
 Q96DG8 PRELIMINARY; PRT; 353 AA.
 AC Q96DG8;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE STIMLAR TO RIKEN CDNA 0610012F22 GENE (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA, AND CHORIOCARCINOMA;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC001563; AH01563.1; -
 FT NON_TER 1
 SQ SEQUENCE 353 AA; 40439 MW; 7B6CE1C847B8D083 CRC64;

Query Match 62.0%; Score 1733; DB 4; Length 353;
 Best Local Similarity 94.6%; Pred. No. 6e-143;
 Matches 334; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

QY 185 TASMLTSPDGPFINLSRLNLTYSQKPKLAKACISISTIMKMLRLMTLALSTESQY 244
 1 TASMLTSPDGPFINLSRLNLTYSQKPKLAKACISISTIMKMLRLMTLALSTESQY 60
 QY 245 KDDMMKVQIGKCYRLGVRARERKQFSALKQOEVDYFLYLAQVYSLDPYALNLFK 304
 61 KDDMMKVQIGKCYRLGVRARERKQFSALKQOEVDYFLYLAQVYSLDPYALNLFK 120
 QY 305 QGLDKPGEVYLLGCIARIYEEMNNSSAEYKVKELKODNTVHXAIACTGSHFYSQPK 364
 121 QGLDKPGEVYLLGCIARIYEEMNNSSAEYKVKELKODNTVHXAIACTGSHFYSQPK 180
 QY 365 EIALRFRRLLOGLYNGQLFNILGICCFYAOQYDMTITSFERALSLAESEEAADWYN 424
 181 EIALRFRRLLOGLYNGQLFNILGICCFYAOQYDMTITSFERALSLAESEEAADWYN 240
 QY 425 LGHVAVGIGDTNLHOCFRALVNNNNHAEVYNNLAVLEMRKGVQEARALLQASSLAP 484
 241 LGHVAVGIGDTNLHOCFRALVNNNNHAEVYNNLAVLEMRKGVQEARALLQASSLAP 300
 QY 485 HMYEPHFNFATISDKIGDLORSYVAAKSEAFPPDHVDTQHLIKOLRQHFAML 537
 301 HMYEPHFNFATISDKIGDLORSYVAAKSEAFPPDHVDTQHLIKOLRQHFAML 353
 DB

RESULT 3
 Q23049
 ID Q23049 PRELIMINARY; PRT; 458 AA.
 AC Q23049;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE T25F10.5 PROTEIN.
 GN T25F10.5
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 NX NCBI_TaxID=6239;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE-94150718; PubMed-7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Riken L., Koopra A., Saunders D., Showkneen R.,
 Smalton N., Smith A., Sonhammer E., Staden K., Sulston J.,
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Pauley, A., Gattung S.;
 RT "The sequence of C. elegans cosmid T25F10.";
 RT Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U64856; AAB04988.1; -
 DR InterPro: IPR000209; Peptidase_58.
 DR InterPro: IPR001440; TPR.
 DR Pfam: PF00515; TPR; 7.
 DR SMART: SM00028; TPR; 4.
 DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
 SQ SEQUENCE 458 AA; 50879 MW; 043203FF5307F7FF CRC64;

RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Abayaratne A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ideyaam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.C.,
 RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Sotler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*.
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rubenstein D.S., McEwen D.G., Pelfer M.A.;
 RT "Drosophila melanogaster O-glycosyltransferase";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003787; AAF57338.1;
 DR EMBL: AF217788; AAF32311.1;
 DR FLYBase: FBgn0040295; Ogt.
 DR InterPro: IPR001440; TPR.
 DR Pfam: PF00515; TPR: 11.
 DR SMART: SM00028; TPR: 11.
 KW transferase.
 SQ SEQUENCE 1059 AA; 118649 MW; 48311D2370321B7D CRC64;

Query Match 6.9%; Score 194; DB 5; Length 1059;
 Best Local Similarity 21.3%; Pred. No. 7.7e-08;
 Matches 68; Conservative 67; Mismatches 140; Indels 44; Gaps 9;

253 IGKYYRLGMYREAEKQFSALK-QQEWDFEVLAKYV-----SLDQPVTL----- 300
 DB 124 LGNFKKRGQLEALDNRRAVRKLPFDITGYNLAALVAARDMESVQAYITLQYNP 183
 QY 301 -----NLFR-----OGIDKFGPEVTLLCGIARIYEMNMSSAAEY 337
 DB 184 DLVCYRSDIGMLKALGLEAEKACYLKAIFTCGFAVAMNLCGVFAQGEIIMLAIHIF 243
 QY 338 KEVIAKQNTHYXALACISNFIYSQPEIALRFYRLLQMGTYNQFLNNIGLCCFYAQ- 396
 DB 244 EKATLDPNFDIATYINLGNVKEARIFEDRAVAALRALNLSPPNNAVVGNDLA-CVYEDQG 302
 QY 397 QYDMLTISFERALSLAEDEEADAVWYNLGHVAVIGDITNLAHQCFRLALVNNNNHAAAY 456

DB 303 LIDLAIDTYRRATELQPN---FPDAYCNLAVALKREKQGVFEAEQCYNTALRLCSNHADSL 359
 QY 457 NNLAVERKRRHVEQARALLQATSSLAAPHMEHPNFATISDKIGDLORS---YVAQKS 513
 DB 360 NNLAINKREQGYIEEARFLYLKALEVFPDFAAHSNLSVLQOQKLEALMHYKEAIRI 419
 QY 514 EAAPFD-HVDQHLIKOLR 531
 DB 420 QPTRDAVSNMGNILKEQ 438

RESULT 8
 ID 026176 PRELIMINARY; PRF: 403 AA.
 AC 026176;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
 DE O-LINKED GLCNAC TRANSFERASE.
 GN MTH72.
 OS Methanothermobacter thermautotrophicus.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanothermobacter.
 CX NCBI_TaxID=145262;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RX MEDLINE-98037514; PubMed-9371463;
 RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pohlier B., Qiu D.,
 RA Spadafora R., Vitare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Yianni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDonnell S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
 RT "Complete genome sequence of *Methanothermobacterium thermautotrophicum*
 RL deltaH: functional analysis and comparative genomics.";
 DR J. Bacteriol. 179:7135-7155(1997).
 DR EMBL: AE000798; AAB84576.1;
 DR InterPro: IPR001440; TPR.
 DR Pfam: PF00515; TPR: 11.
 DR SMART: SM00028; TPR: 11.
 KW transferase; Complete proteome.
 SQ SEQUENCE 403 AA; 46660 MW; F3539389ED713613 CRC64;

Query Match 6.6%; Score 186; DB 17; Length 403;
 Best Local Similarity 22.1%; Pred. No. 9.3e-08;
 Matches 72; Conservative 49; Mismatches 147; Indels 58; Gaps 7;

245 KDW-----WNRQVIGKCYRLGMYREAEKQFSALKQQEWDFVL-YLAKYVSLDQPV 297
 DB 8 KDWIRAGKAGWHLISGGRSSLKQGYKFKALKSPNDPEILHNAMTLTKRHP 67
 QY 298 TALNIFKQGLDKFPGPEVTLLCGIARIYEMNMSSAAEYKEVLKOD-----NTHYXA 350
 DB 68 KALKCYERTLKNPNPLAEAMNNGVLELKRKYDALCYERALQIDPDQDGTWNNKGAL 127
 QY 351 IACISNFIYSQPEIALRFYRLLQMGTYNQFLNNIG-----LCCP----- 393
 DB 128 LDTIG-----KPKALIECYEKALEINOKNAKAYNNGNGLSGKYEALIECYEKALQ 180
 QY 394 -----YAQYDMLTISFERALSLAEDEEADAVWYNLGHVAVIGDITNL 436
 DB 181 INAEFEVAMYNKALIFEELEKRYDEALECYERALQIDPDQDGTWNNKGAL 237
 QY 437 LAHQCFRLALVNNNNHAAAYNVLAVLEMRKGVHEDARALLQYASSLAAPHMEHPNFATI 496
 DB 238 KALIECYEKALEINOKNAAMNNGVLEELKRYDEALECYEKALEINLENDETWANKVYL 297
 QY 497 SDKIGDLORSYVAQKSEAFPDHVD 522

Db 298 LRLKGYEALCEFEKALEINPEPAD 323

RESULT 9

096H45 PRELIMINARY; PRT: 464 AA.

ID 096H45 PRELIMINARY; PRT: 464 AA.

AC 096H45;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE UNKNOWN (PROTEIN FOR IMAGE:3659371) (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-ENDOMETRIAL ADENOCARCINOMA;

RA Strausberg R.;

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC008923; AA08923.1; -

FT NON_TER

FT 1 1

SEQUENCE 464 AA: 51674 MW; 32824ADDFE3E3420 CRC64;

Query Match 6.6%; Score 185.5; DB 4; Length 464;

Best Local Similarity 24.5%; Pred. No. 1.3e-07;

Matches 67; Conservative 50; Mismatches 147; Indels 9; Gaps 3;

Qy 245 KDMKKKVOIGKCYRLGMYREAEKQFSAKLNQOEMVDFPLYLAKYV---SLDQPYTALN 301

Db 78 KMEIISHNIGVCYIYLKQKNAQDQHNALNLRHDLTYIMLGKTHLEGDLK--ALE 134

Qy 302 LEKQGLDKFPGFVETLLCGIARIYEEMNNSSAAEYKEVLEKODNTHVXAIACISNHFYS 361

Db 135 VYKKAVERSPETELLTTLGLIYLQIGYKAFELHGNALTYDPRNYKAILLAGSMQTH 194

Qy 362 DQPEALRFRRLLDMGTYNGOLFNNLGLCCFYAQOYDMTLTSFERALSIAENEEDADV 421

Db 195 GDFVALTKRYRVACAVESPPLMNNIGMCFGKRYVAALISCLRANLAFPMK---I 251

Qy 422 WNLGHVAVGIDGTLNLAQCFRLALVNNNNHAEAVNNLAVLEMRGHVQAARALLOTASS 481

Db 252 LYNLGLVHLTMQOYASARHFLSALINFPQKMGELTMLLAVALTNIJEDIENARAYAEVH 311

Qy 482 LAPHMYEPHFNFATISDRKIGDLORSYVAAQKSE 514

Db 312 LDKCNPLVNLNVAVLLYNGEKKNALAQYQEME 344

RESULT 10

096CC1 PRELIMINARY; PRT: 1036 AA.

ID 096CC1 PRELIMINARY; PRT: 1036 AA.

AC 096CC1;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE O-LINKED N-ACETYLGLUCOSAMINE (GLCNAC) TRANSFERASE

DE (UDP-N-ACETYLGLUCOSAMINE:POLYPEPTIDE-N-ACETYLGLUCOSAMINYL TRANSFERASE)

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-COLON ADENOCARCINOMA;

RA Strausberg R.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC014434; AA014434.1; -

FT TRANSFERASE.

FT 1036 AA: 115705 MW; C3BD67340925A2C2 CRC64;

Query Match 6.6%; Score 185.5; DB 4; Length 1036;

Best Local Similarity 22.0%; Pred. No. 4.1e-07;

Matches 68; Conservative 53; Mismatches 139; Indels 49; Gaps 8;

Qy 253 IGKCYRLGMYREAEKQFSAKLNQOEMVDFPLYLAKYVSLDQPYTALNFKQGLDKFP 311

Db 85 LGNVYKERQGLQDAIEHYHNRALRKPFDIGYINLAALVYAGDMGVAQAAVSALQYNP 144

Qy 312 GCVTLTLCGIARIYEEMNNSSAAEYKEVLEKODNTHVXAIACISNHFYSQDPEI---- 366

Db 145 DLICVRSDLGNLKLALGRLEAKACTYKAIETQPN---FAVMSNLGCVFNAGGEIMLAI 201

Qy 367 -----ALREFRLLDMGTYNGOLFNNLGLCCFY 394

Db 202 HHFEKAVTLDPNFDLAYINLGNVLEKARIFEDRAVAAYLRALSLSPNHAVHGNLA-CVY 260

Qy 395 AQ-QYDMTLTSFERALSIAENEEDADVYNIQHAVGIGDTNLAHQCFRLALVNNNNHA 453

Db 261 EQLGLIDLADITYRAIEL---QHPPDAYCNLANLKEKGSVAEEDCYNLRILCPPTHA 317

Qy 454 EAYNNULAVLEMRGHVQAARALLOTASSLAPHMYEPHFNFATISDRKIGLORS---YVAA 510

Db 318 DSLNLANLRKQGNIEEAVRLYRKALVEFPFPAASHSLASVLOQOGKLODALMHRKA 377

Qy 511 OKSEAAFPD 519

Db 378 IRISPTFAD 386

RESULT 11

091Y38 PRELIMINARY; PRT: 1046 AA.

ID 091Y38;

AC 091Y38;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE UDP-N-ACETYLGLUCOSAMINYLTRANSFERASE.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BALB/C; TISSUE-SKELETAL MUSCLE;

RA Rumberger J.M., Wu T., Hering M.A., Marshall S.;

RL "Molecular cloning of the mouse O-GlcNAc transferase.";

RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF363030; AAK39123.1; -

FT Transferase; Glycosyltransferase.

FT 1046 AA: 116979 MW; 0D47F82844E0130A CRC64;

SEQUENCE 1046 AA: 116979 MW; 0D47F82844E0130A CRC64;

Query Match 6.6%; Score 185.5; DB 11; Length 1046;

Best Local Similarity 22.0%; Pred. No. 4.2e-07;

Matches 68; Conservative 53; Mismatches 139; Indels 49; Gaps 8;

Qy 253 IGKCYRLGMYREAEKQFSAKLNQOEMVDFPLYLAKYVSLDQPYTALNFKQGLDKFP 311

Db 95 LGNVYKERQGLQDAIEHYHNRALRKPFDIGYINLAALVYAGDMGVAQAAVSALQYNP 154

Qy 312 GCVTLTLCGIARIYEEMNNSSAAEYKEVLEKODNTHVXAIACISNHFYSQDPEI---- 366

Db 155 DLICVRSDLGNLKLALGRLEAKACTYKAIETQPN---FAVMSNLGCVFNAGGEIMLAI 211

Qy 367 -----ALREFRLLDMGTYNGOLFNNLGLCCFY 394

Db 212 HHFEKAVTLDPNFDLAYINLGNVLEKARIFEDRAVAAYLRALSLSPNHAVHGNLA-CVY 270

Qy 395 AQ-QYDMTLTSFERALSIAENEEDADVYNIQHAVGIGDTNLAHQCFRLALVNNNNHA 453

Db 271 EQLGLIDLADITYRAIEL---QHPPDAYCNLANLKEKGSVAEEDCYNLRILCPPTHA 327

D 237 GDLNRALQYKEAVKLRFPDPAVLNIGNVKALGRPEALIMCQHALQMPNSAMAEGN 296
Q 320 IARIYEENNNNSAAEYKEVILKODNTHVXALICIGSNHFPSSDOPELAFRYRRLQMGCI 379
D 297 IASITIEGGQDLAIIRHKQALSRDPRLEAVNNLGNALIKIGRDEAVRCYNOCLOAP 356
Q 380 YNGOLFNNLG-----LCCFYAQO--YDMTITNSF 405
D 357 NHPOAMNLGNIYMEWNMMGPASSLFRKATLAVTTGLSAPFNNLAIITYKQCNYSDAISCY 416
Q 406 ERAISLAENEEDADWYNLGHVAVGIDTILAQCFPLALVNNNNHAEAVNNLAVLEMR 465
D 417 NEVARI--DPLADALVNRGNTYKEIGRYTEALIDYMHAINFRPTMAEHAHLASAYKD 473
Q 466 KGHVEQARALLQTAASSLAPHYEPHFH 492
D 474 SGHVEAATTSYKQALLLRDPPEATCN 500

RESULT 14
Q96RR4 PRELIMINARY; PRT; 519 AA.
AC 096RR4
DT 01-DEC-2001 (TREMblrel, 19, Created)
DT 01-DEC-2001 (TREMblrel, 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel, 19, Last annotation update)
DE BARDE-BIEDL SYNDROME TYPE 4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21275976; PubMed=11381270;
RA Mykyryn K., Braun T., Carni R., Haider N.B., Seaby C.C., Shastri M.,
RA Beck G., Wright A.F., Iannaccone A., Elbedour K., Riise R., Baldi A.,
RA Raas-Nielsen A., Gorman S.W., Duhl D.M., Jacobson S.G.,
RA Casavant T., Stone E.M., Sheffield V.C.;
RT "Identification of the gene that, when mutated, causes the human
RT obesity syndrome BBS4".
RT Mt. Genet. 28:188-191(2001).
DR EMBL: AF359281; AAK58868.1;
SQ SEQUENCE 519 AA; 58270 MW; B3AA2232BE7F530A CRC64;

Query Match 6.5%; Score 182.5; DB 4; Length 519;
Best Local Similarity 24.5%; Pred. No. 2.7e-07;
Matches 67; Conservative 49; Mismatches 148; Indels 9; Gaps 3;

Q 245 KDMWVKVQIGKCYRYLGMVREAEKQFSALKQOEWDVDFLAKVYV--SLDQPVTA LN 301
D 133 KDWELSHNLGVCYIYLKQFNKAQODLNHALNLNRHDLIYIMGRKHLLEGSLDK---AIE 189
Q 302 LFKGGLDKFGEVTLGCIARIYEENNNNSAAEYKEVILKODNTHVXALICIGSNHFPYS 361
D 190 YKKRAVESPEPNTLLTTLTGLLYLQIGYKAFHGLNALTYPNTYKAILAAGSMQGTJ 249
Q 362 DQPELALREYERLQMGVINGOLFNNLGLCFYAOQYMTLITSPERALSIAENEEDADV 421
D 250 GDFVVALTKRVACAVPESPRLNNNIGMCFGRKKYVAALSCUKRAVYLA PFDWK--I 306
Q 422 WYNLGHVAVGIDTILAQCFPLALVNNNNHAEAVNNLAVLEMRKGHVEQARALLQTAASS 481
D 307 LYNLGLVHLTMQOYASAFHLSAALINFPKMGELIYMLALVLTNLEDTENAKRAYAEAVH 366
Q 482 LAPHYEPHFATISDKIGLQSRSYAAQKSE 514
D 367 LDKCNPLVNLNVAVLTYNKGKKNALAOYQEME 399

RESULT 15
O18158

ID O18158 PRELIMINARY; PRT; 1151 AA.
AC O18158; Q21232;
DT 01-NOV-1998 (TREMblrel, 08, Created)
DT 01-NOV-1998 (TREMblrel, 08, Last sequence update)
DT 01-JUN-2001 (TREMblrel, 17, Last annotation update)
DE UDP-N-ACETYLGLUCOSAMINE-PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE
DE (EC 2.4.1.1) (O-GLCNAC) (OGT).
GN K04G7.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 321-340 AND 1060-1076.
RX MEDLINE=97238870; PubMed=9083068;
RA Lubas W.A., Frank D.W., Krause M., Hancock J.A.;
RT "O-linked GlcNAc transferase is a conserved nucleocytoplasmic protein
RT containing tetratricopeptide repeats.";
RL J. Biol. Chem. 272:9316-9324(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berts M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kersey J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonnenhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Welnslock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- FUNCTION: ADDITION OF NUCLEOTIDE-ACTIVATED SUGARS DIRECTLY ONTO
CC THE POLYPEPTIDE THROUGH O-GLYCOSIDIC LINKAGE WITH THE HYDROXYL OF
CC SERINE OR THREONINE.
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + PEPTIDE -> UDP +
CC N-ACETYL-BETA-D-GLUCOSAMINYL-PEPTIDE.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC (POSSIBLE).
CC -1- SIMILARITY: CONTAINS 12 COPIES OF THE TPR DOMAIN.
DR EMBL: U77412; AAB63465.1;
DR EMBL: U21320; AAA62535.1; ALT_SEQ.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR. 12.
DR SMART: SM00028; TPR. 11.
KW Transferase; Glycosyltransferase; Nuclear protein; Repeat;
KW TPR domain.
FT REPEAT 193 226 TPR 1.
FT REPEAT 227 260 TPR 2.
FT REPEAT 261 294 TPR 3.
FT REPEAT 295 328 TPR 4.
FT REPEAT 329 362 TPR 5.
FT REPEAT 363 396 TPR 6.
FT REPEAT 397 430 TPR 7.
FT REPEAT 431 464 TPR 8.
FT REPEAT 465 498 TPR 9.
FT REPEAT 499 532 TPR 10.
FT REPEAT 533 566 TPR 11.
FT REPEAT 567 577 TPR 12 (INCOMPLETE).
FT DOMAIN 591 607 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 1151 AA; 127981 MW; DIC8F69401CC53B CRC64;

Query Match 6.4%; Score 178; DB 5; Length 1151;
Best Local Similarity 22.8%; Pred. No. 2.2e-06;
Matches 104; Conservative 73; Mismatches 195; Indels 84; Gaps 19;

Q 129 QAVRPTQAGRPTIGFLRPSQ---SGRPQTEQAIPTPTAYTARPTSSGRRVRL- 184

```

Db 64 QAIR--TQLENOAAQOLAVNQOFQNGATAVQOQLLTPOOS-LAOPIALAPQPTVYVNG 120
QY 185 TASMUTS-----PDGPFINLSR-LNLTYSQKPKLAKACG-SISFTMKMLR-LL 231
Db 121 VSETLKVAELAHROFQSGNVEAEKYNLFQSDPNNLPTLLLSAINFQTKNLEKSMQ 180
QY 232 WIMWALSTESHOYKIDMMWKVOIGKCYRIGMYREAEKQFKSALK-----QOEMVDTF 283
Db 181 YSMALAIKVN-----QCAEAYSNLGNYEKEGLODLENYKLAVKLKPEFTDAY 230
QY 284 LYIAKYVS--IDQPYTALNLFKQGLDKPEGEVTLGIGARIYEEMNNMSSAAEYXKEV 340
Db 231 INLAALVSGDLEQAVTA--YFNALQINPDLYCVRSDLGNLTKAMGRLEAKVCYLKA 287
QY 341 LKQDNTHVXAIACIGSNHFSQPEI--ALRFYRRLQMGITYNQOLFNNLGLCCFYAQY 398
Db 288 IETQPOFAVAMWSNLGC--VENSQGEIMLAIHFEKAVTLDPNPLDAYINLGNVLKEARIF 345
QY 399 DMTLTSFERALSLAENEE---EAADYWNLGHVAVGIGDT-----NLA 438
Db 346 DRAVSATLRALNLSGNHAYVHGNLACYIEQGLIDLAI-DTYKKAIDLQHPHPPDAYCNLA 404
QY 439 H-----OCFRIALVNNNNHAEAYNNLAVLEMRKGHVQARALLQFASSLAPHM 486
Db 405 NALKEKGSVVEAEQMYKMALELCPTHADSONNLANIKREGKIEDATRLYLKALEITYPEF 464
QY 487 YEPHFNFATF--SDKIGDLORSYVAAQKSEAAPPD 519
Db 465 AAHSNLSASIIQQQGLNDAILHYKEAIRIAPTFAD 500

```

Search Completed: July 17, 2002, 16:45:49
 Job time: 145 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 16:46:15 ; Search time 13.48 Seconds

(without alignments)
1542.463 Million cell updates/sec

Title: US-09-821-812-5

Sequence: 1 GGLTTTVIGTRLGVDRPRLS.....PDHVDTOHLIKLRQHFAML 537

Scoring table: OLIGO

Gapop 60.0 , Gapept 60.0

Searched: 105224 seqs, 38719550 residues

Word size: 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.5	293	1	VI99_AGRG9
2	8	1.5	532	1	MURD_TREPA
3	8	1.5	1395	1	CUT1_MOUSE
4	7	1.3	74	1	RS18_ECOLI
5	7	1.3	74	1	RS18_HAEIN
6	7	1.3	75	1	RS18_BUCAT
7	7	1.3	75	1	RS18_PASMU
8	7	1.3	106	1	BPY2_HUMAN
9	7	1.3	110	1	NOLE_RHILP
10	7	1.3	133	1	Y044_BORBU
11	7	1.3	137	1	RS9_SURSO
12	7	1.3	137	1	RS9_SURTO
13	7	1.3	143	1	PER_DROFI
14	7	1.3	147	1	YS72_MYCTU
15	7	1.3	174	1	YC52_PORPU
16	7	1.3	215	1	ALU7_DERPT
17	7	1.3	223	1	GL1A_ARATH
18	7	1.3	223	1	GL1B_ARATH
19	7	1.3	229	1	LOLD_BUCAP
20	7	1.3	272	1	VB08_VACCC
21	7	1.3	272	1	VB08_VACCV
22	7	1.3	281	1	VPM_BPHP1
23	7	1.3	285	1	FMKB_ECOLI
24	7	1.3	288	1	BSN2_BACSU
25	7	1.3	292	1	TF_RABIT
26	7	1.3	315	1	T2A1_ANASP
27	7	1.3	316	1	KHSE_PSEAE
28	7	1.3	319	1	METX_MERTH
29	7	1.3	330	1	V258_HAEIN
30	7	1.3	340	1	AC11_HUMAN
31	7	1.3	341	1	AN11-COLLI
32	7	1.3	343	1	AN12-COLLI
33	7	1.3	373	1	YE52_LISMO

34	7	1.3	373	1	YE89_LISIN	O92bq8	listeria in
35	7	1.3	374	1	ALFC_CHIRE	O42650	chlamydomon
36	7	1.3	384	1	GLF1_KLEPE	O48485	klebsiella
37	7	1.3	385	1	BIOF_ERWHE	O47829	erwinia her
38	7	1.3	395	1	TRBL_AGRU	P54913	agrobacteri
39	7	1.3	412	1	CLPX_AQJAE	O67356	aquilex aeo
40	7	1.3	420	1	CLPX_CAUCR	O67708	caulobacter
41	7	1.3	423	1	CLPX_ECOLI	P33138	escherichia
42	7	1.3	423	1	CLPX_YEREN	O33873	yersinia en
43	7	1.3	427	1	CLPX_MTXAA	O935n1	myxococcus
44	7	1.3	429	1	CLPX_BUCAT	P57548	buchnera ap
45	7	1.3	488	1	DHAL_PSESP	P33008	pseudomonas

ALIGNMENTS

```

RESULT 1
VI99_AGRG9 STANDARD; PRT; 293 AA.
AC P05358; P09782;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE VI99 protein precursor.
GN VI99.
OS Agrobacterium tumefaciens.
OC Plasmid pTI15955, and plasmid pTIAG.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_Taxid=358;
RN [1]
RP SEQUENCE FROM N.A.
RC PLASMID-P115955;
RX MEDLINE=88247765; PubMed=2837739;
RA Thompson D.V., Melchers L.S., Idler K.B., Shilperoot R.A.,
RA Hooykaas P.J.J.;
RT "Analysis of the complete nucleotide sequence of the Agrobacterium
RT tumefaciens virB operon."
RL Nucleic Acids Res. 16:4621-4636(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID-P115955;
RX MEDLINE=88166901; PubMed=3281947;
RA Ward J.E., Akiyoshi D.E., Regier D., Datta A., Gordon M.P.,
RA Nester E.W.;
RT "Characterization of the virB operon from an Agrobacterium
RT tumefaciens Ti plasmid."
RL J. Biol. Chem. 263:5804-5814(1988).
RN [3]
RP REVISIONS.
RX MEDLINE=90170994; PubMed=2307685;
RA Ward J.E., Akiyoshi D.E., Regier D., Datta A., Gordon M.P.,
RA Nester E.W.;
RL J. Biol. Chem. 265:4768-4768(1990).
-!- FUNCTION: VIRB PROTEINS ARE SUGGESTED TO ACT AT THE BACTERIAL
SURFACE AND THERE PLAY AN IMPORTANT ROLE IN DIRECTING T-DNA
TRANSFER TO PLANT CELLS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X06826; CAA29979.1; -
CC DR EMBL: J03216; AAA88654.1; -
CC DR PIR: S00785; B9AG55.
CC PIR: B30402; B9AG46.
CC Crown gall tumor; Plasmid; Signal.
CC SIGNAL 21 POTENTIAL.

```

FT CHAIN 22 293 VIRB9 PROTEIN.
SQ SEQUENCE 293 AA; 32172 MW; 1BA2E48058E7DD19 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 293;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 LDQPYTAL 300
Db 173 LDQPYTAL 180

RESULT 2
MURD_TREPA STANDARD; PRT; 532 AA.
ID MURD_TREPA
AC 083873;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylmuramoylalanine-D-glutamate ligase (EC 6.3.2.9) (UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase) (D-glutamic acid adding enzyme).
DE MURD OR TP0903.
GN Treponema pallidum.
OS Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OC NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOIS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G., Dodson R., Glyn M., Hickey E.K., Clayton R., Ketchum K.A., Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J., Khailak L., Richardson D., Howell J.K., Chidambaram M., Uetebach T., McDonald H., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S., Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O., Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis spirochete."
RL Science 281:375-388(1998).
CC -1- FUNCTION: CELL WALL FORMATION. CATALYZES THE ADDITION OF D-GLUTAMATE TO THE NUCLEOTIDE PRECURSOR UDP-N-ACETYLURAMOYL-L-ALANINE (UMA) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanine + D-glutamate -> ADP + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamate.
CC -1- PATHWAY: PEPTIDOLYCAN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO THE MURDEF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001259; AAC65856.1;
DR HSSP: P14800; IEOD.
DR TIGR: TP0903;
DR InterPro: IPR000713; Mur_ligase.
DR InterPro: IPR004101; Mur_ligase_C.
DR Pfam: PF01225; Mur_ligase_1.
DR Pfam: PF02875; Mur_ligase_C_1.
KM Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
KW ATP-binding; Complete proteome.
SQ NP_BIND 124 130 ATP (POTENTIAL).
FT SEQUENCE 532 AA; 57847 MW; 9FD705C2AE478ADD CRC64;

Query Match 1.5%; Score 8; DB 1; Length 532;

Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 EORALLQ 477
Db 2 EORALLQ 9

RESULT 3
CUTL_MOUSE STANDARD; PRT; 1395 AA.
ID CUTL_MOUSE
AC P53564; 008994; P70301;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CCAAT displacement protein (CDP) (Cut-like 1) (Homeobox protein Cux) (Fragment).
DE CUTLI OR CUX.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6N;
RA Wang Z., Goldstein A., Neufeld E.J., Scheuermann R.H., Tucker P.W.;
RT "Repression of immunoglobulin heavy chain intronic enhancer through nuclear matrix attachment sites: Cux/CDP homeoprotein is a component of NF-muR repressor."
RL Submitted (May-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 64-1395 FROM N.A. (ISOFORM 1).
RC STRAIN=A/J, AND BALB/C; TISSUE=Brain;
RX MEDLINE=94244481; PubMed=7910552;
RA Valarche I., Tisler-Seta J.P., Hirsch M.R., Martinez S., Goridis C., Brunet J.F.;
RT "The mouse homeodomain protein Pbx2 regulates Ncam promoter activity in concert with Cux/CDP and is a putative determinant of neurotransmitter phenotype."
RL Development 119:881-896(1993).
RN [3]
RP SEQUENCE OF 642-1395 FROM N.A.
RX MEDLINE=96437626; PubMed=8840273;
RA den Heuvel G.B., Bodmer R., McConnell K.R., Nagami G.T., Igarashi P.;
RT "Expression of a cut-related homeobox gene in developing and polycystic mouse kidney."
RL Kidney Int. 50:453-461(1996).
RN [4]
RP SEQUENCE OF 936-1395 FROM N.A.
RC TISSUE=Testis;
RA Quaglin S.E., Igarashi P.;
RT "A unique variant of a homeobox gene related to Drosophila cut is expressed in mouse testis."
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY HAS A BROAD ROLE IN MAMMALIAN DEVELOPMENT AS A REPRESSOR OF DEVELOPMENTALLY REGULATED GENE EXPRESSION. MAY ACT BY PREVENTING BINDING OF POSITIVELY-ACTIVATING CCAAT FACTORS TO PROMOTERS (BY SIMILARITY). COMPONENT OF NF-MUR REPRESSOR; BINDS TO THE MARS (5' AND 3') OF THE IMMUNOGLOBULIN HEAVY CHAIN ENHANCER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 3 CUT DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE CUT FAMILY OF HOMEOBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

CC EMBL: AF004225; AAD12485.1; -
 DR EMBL: X75013; AAC52922.1; -
 DR EMBL: U46683; AAC52775.1; -
 DR EMBL: U46684; AAB41146.1; -
 DR HSSP: P10037; IAD7.
 DR MGD: MGI:88568; Cui1.
 DR InterPro: IPR003350; Cui.
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF02376; Cui; 3.
 DR Pfam: PF00046; homeobox; 2.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR Transcription regulation; Homeobox; DNA-binding;
 KW Developmental protein; Nuclear protein; Repeat; Repressor;
 KW Coiled coil; Alternative splicing.
 FT NON_TER 1 1
 FT DOMAIN 1 243 COILED COIL (POTENTIAL).
 FT DNA_BIND 420 507 CUT 1.
 FT DOMAIN 547 603 COILED COIL (POTENTIAL).
 FT DNA_BIND 809 896 CUT 2.
 FT DNA_BIND 992 1079 CUT 3.
 FT DNA_BIND 1119 1178 HOMEBOX.
 FT VARSPLIC 287 388 MISSING (IN ISOFORM 2).
 FT CONFLICT 1360 1360 G -> A (IN REF. 2).
 FT CONFLICT 1365 1365 P -> L (IN REF. 1).
 SQ SEQUENCE 1395 AA; 151802 MW; D062CC227D7A1636 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 1395;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 PSIAAPAA 32
 Db 1293 PSIAAPAA 1300

RESULT 4
 ID RS18_ECOLI STANDARD; PRT; 74 AA.
 AC P02374;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 30S ribosomal protein S18.
 GN RPSR OR B4202 OR Z5811 OR ECS5178 OR STY4749.
 OS Escherichia coli,
 OS Escherichia coli O157:H7, and
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334, 601;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP SPECIES=E.coli;
 RC MEDLINE=86310297; PubMed=3528756;
 RA Schnier J., Kitakawa M., Isono K.;
 RT "The nucleotide sequence of an Escherichia coli chromosomal region
 RT containing the genes for ribosomal proteins S6, S18, L9 and an open
 RT reading frame";
 RL Mol. Gen. Genet. 204:126-132(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;
 RX MEDLINE=95334362; PubMed=7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [3]

RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potomous K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509592;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shida T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 RN [5]
 RP SEQUENCE.
 RC SPECIES=E.coli; STRAIN=K;
 RX MEDLINE=76210737; PubMed=776663;
 RT Yaguchi M.;
 RT "Primary structure of protein S18 from the small Escherichia coli
 RT ribosomal subunit";
 RL FEBS Lett. 59:217-220(1975).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhi; STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Beltrami T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18";
 RL Nature 413:848-852(2001).
 RN [7]
 RP MASS SPECTROMETRY.
 RC SPECIES=E.coli;
 RX MEDLINE=99196679; PubMed=10094780;
 RA Arnold R.J., Bellly J.P.;
 RT "Observation of Escherichia coli ribosomal proteins and their
 RT posttranslational modifications by mass spectrometry";
 RL Anal. Biochem. 269:105-112(1999).
 CC CC -1- FUNCTION: THIS PROTEIN HAS BEEN IMPLICATED IN AMINOACYL-TRANSFER
 CC RNA BINDING. IT APPEARS TO BE SITUATED AT THE DECODING SITE OF
 CC MESSENGER RNA.
 CC CC -1- MASS SPECTROMETRY: MW=8897.0; METHOD=MALDI.
 CC CC -1- SIMILARITY: BELONGS TO THE S18P FAMILY OF RIBOSOMAL PROTEINS.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X04022; CA27654.1; -
 DR EMBL: U14003; AAA97098.1; -
 DR EMBL: AE000491; AAC77159.1; -
 DR EMBL: AE000542; AAG59398.1; -
 DR EMBL: AP002568; BAB38601.1; -
 DR EMBL: AL627283; CAD06870.1; -

DR PIR: A02741; R3EC18.
 DR EcoGene; EG10917; tpsr.
 DR InterPro; IPR001648; Ribosomal_S18.
 DR Pfam; PF01084; Ribosomal_S18; 1.
 DR PRINTS; PR00974; RIBOSOMALS18.
 DR PRODOM; PD002239; Ribosomal_S18; 1.
 DR PROSITE; PS00057; RIBOSOMAL_S18; 1.
 KW Ribosomal protein; Acetylation; RNA-binding; Complete proteome.
 FT INIT_MET 0
 FT MOD_RES 1
 FT CONFLICT 15 E -> Q (IN REF. 5).
 FT CONFLICT 23 K -> E (IN REF. 3).
 SO SEQUENCE 74 AA; 8855 MW; AB2EDDEA9441581 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 74;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 46 YFRRKRF 52
 |||||||
 Db 3 YFRRKRF 9

RESULT 5
 RS18_HAEIN STANDARD; PRT; 74 AA.
 AC P44384;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 30S ribosomal protein S18.
 GN RPSR OR RPS18 OR H10545.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 OX NCBI_TaxId=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kelleys A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
 RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodex A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Grehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT Influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: THIS PROTEIN HAS BEEN IMPLICATED IN AMINOACYL-TRANSFER
 CC RNA BINDING. IT APPEARS TO BE SITUATED AT THE DECODING SITE OF
 CC MESSENGER RNA (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S18P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL; U03736; AAC22203.1; -;
 DR TIGR; H10545; -;
 DR InterPro; IPR001648; Ribosomal_S18.
 DR Pfam; PF01084; Ribosomal_S18; 1.
 DR PRINTS; PR00974; RIBOSOMALS18.
 DR PRODOM; PD002239; Ribosomal_S18; 1.
 DR PROSITE; PS00057; RIBOSOMAL_S18; 1.
 DR PROSITE; PS00057; RIBOSOMAL_S18; 1.

KW Ribosomal protein; Acetylation; RNA-binding; Complete proteome.
 FT INIT_MET 0
 FT MOD_RES 1
 FT CONFLICT 15 E -> Q (IN REF. 5).
 FT CONFLICT 23 K -> E (IN REF. 3).
 SO SEQUENCE 74 AA; 8811 MW; AC39A87BC54D1C81 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 74;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 46 YFRRKRF 52
 |||||||
 Db 3 YFRRKRF 9

RESULT 6
 RS18_BUCAI STANDARD; PRT; 75 AA.
 AC P57626;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 30S ribosomal protein S18.
 GN RPSR OR B0563.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 OS symbiotic bacterium).
 CC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxId=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TOKYO 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 RT Buchnera sp. Aps.";
 RL Nature 407:81-86(2000).
 CC -1- FUNCTION: THIS PROTEIN HAS BEEN IMPLICATED IN AMINOACYL-TRANSFER
 CC RNA BINDING. IT APPEARS TO BE SITUATED AT THE DECODING SITE OF
 CC MESSENGER RNA (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S18P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL; AP001119; BAB13253.1; -;
 DR InterPro; IPR001648; Ribosomal_S18.
 DR Pfam; PF01084; Ribosomal_S18; 1.
 DR PRINTS; PR00974; RIBOSOMALS18.
 DR PRODOM; PD002239; Ribosomal_S18; 1.
 DR PROSITE; PS00057; RIBOSOMAL_S18; 1.
 KW Ribosomal protein; RNA-binding; Complete proteome.
 SO SEQUENCE 75 AA; 8984 MW; 445B82A691FB73BA CRC64;

Query Match 1.3%; Score 7; DB 1; Length 75;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 46 YFRRKRF 52
 |||||||
 Db 4 YFRRKRF 10

RESULT 7
 RS18_PASMU STANDARD; PRT; 75 AA.
 AC P57916;
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE 30S ribosomal protein S18.
 GN RPSR OR RPS18 OR Pml178.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 NCBI_TaxId=747;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida pm70.";
 Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001)
 CC -1- FUNCTION: THIS PROTEIN HAS BEEN IMPLICATED IN AMINOACYL-TRANSFER
 RNA BINDING. IT APPEARS TO BE SITUATED AT THE DECODING SITE OF
 CC MESSENGER RNA (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S18P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE006158; AK03262.1; -
 DR InterPro; IPR001648; Ribosomal_S18.
 DR Pfam; PF01084; Ribosomal_S18; 1.
 DR PRINTS; PR00974; RIBOSOMALS18.
 DR ProDom; PD002239; Ribosomal_S18; 1.
 DR PROSITE; PS00057; RIBOSOMAL_S18; 1.
 KW Ribosomal protein; RNA-binding; Complete proteome.
 SQ SEQUENCE 75 AA; 8942 MW; AC284AAB36C27AB8 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 75;
 Best Local Similarity 100.0%; Pred. No. 9.8; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 YFRRRKF 52
 |||||
 DB 4 YFRRRKF 10

RESULT 8
 BPY2_HUMAN
 ID BPY2_HUMAN STANDARD; PRT; 106 AA.
 AC O14599;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Testis-specific basic protein Y 2.
 GN BPY2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxId=9606;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98022381; PubMed=9381176;
 RA Lahn B.T., Page D.C.;
 RT "Functional coherence of the human Y chromosome.";
 Science 278:675-680(1997).
 CC -1- TISSUE SPECIFICITY: TESTIS-SPECIFIC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF000980; AAC51828.1; -
 DR MIM; 400013; -
 SQ SEQUENCE 106 AA; 12035 MW; 2573EC02A2DEA788 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 106;
 Best Local Similarity 100.0%; Pred. No. 13; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 PGEVTL 317
 |||||
 DB 77 PGEVTL 83

RESULT 9
 NOLE_RHLP
 ID NOLE_RHLP STANDARD; PRT; 110 AA.
 AC P23716;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Nodulation protein NOLE precursor.
 GN NOLE.
 OS Rhizobium leguminosarum (biovar phaseol).
 OC Plasmid sym.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxId=385;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=8002;
 RX MEDLINE=91014692; PubMed=2215216;
 RA Davis E.O., Johnston A.W.B.;
 RT "Analysis of three nod genes in Rhizobium leguminosarum biovar
 RT phaseol: nodD is preceded by nole, a gene whose product is secreted
 RT from the cytoplasm.";
 RL Mol. Microbiol. 4:921-932(1990).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X54214; CA38125.1; -
 DR PIR; S11786; S11786.
 DR Fm Plasmid; Nodulation; Periplasmic; Signal.
 KW SIGNAL 1 POTENTIAL.
 FT CHAIN 26 110 NODULATION PROTEIN NOLE.
 SQ SEQUENCE 110 AA; 12050 MW; EF2FB42A7BDC904 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 14; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 AGPSIAA 29
 |||||
 DB 20 AGPSIAA 26

RESULT 10
 Y044_BORBU
 ID Y044_BORBU STANDARD; PRT; 133 AA.
 AC O51073;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein BB0044.
 GN BB0044.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 CC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 CX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Ladhara R., White O., Kechum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kellavag A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Usterback T., Matthey L., McDonald L., Artlich P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.";
 RL Nature 390:580-586(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL; AE001118; AAC66442.1; -;
 DR TIGR; BB0044; -;
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 11 31 POTENTIAL.
 SQ SEQUENCE 133 AA; 16052 MW; A57686EA30F1959B CRC64;

Query Match 1.3%; Score 7; DB 1; Length 133;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 TOMLEKS 66
 Db 31 TOMLEKS 37

RESULT 11
 RS9_SULTSO STANDARD; PRT: 137 AA.
 AC P95992;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 30S ribosomal protein S9P
 GN RPS9P OR RPS9AB OR RPS9 OR SS00068 OR C05003.
 OS Sulfolobus solfataricus.
 CC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 CX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=97053432; PubMed=8899719;
 RA Sersen C.W., Klenk H.-P., Singh R.K., Allard G., Chan C.C.-Y.,
 RA Liu Q.-Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
 RA Doolittle W.F., Ragan M.A., Charlebois R.L.;
 RT "Organizational characteristics and information content of an
 RT archaeal genome: 156 kb of sequence from Sulfolobus solfataricus
 RT P2.";
 RL MOL. Microbiol. 22:175-191(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;

RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aweyer M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sersen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 CC -1- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL; Y08257; CAA69534.1; ALT_INIT.
 DR EMBL; AE006647; AAK0430.1; ALT_INIT.
 DR InterPro: IPR000754; Ribosomal_S9.
 DR Pfam: PF00380; Ribosomal_S9; 1.
 DR ProDom: PD001627; Ribosomal_S9; 1.
 DR PROSITE: PS00360; RIBOSOMAL_S9; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 137 AA; 15705 MW; 127BAE7C72F8B410 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 137;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 MEPLILA 43
 Db 50 MEPLILA 56

RESULT 12
 RS9_SULTSO STANDARD; PRT: 137 AA.
 AC O96YK3;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 30S ribosomal protein S9P
 GN RPS9P OR ST2064.
 OS Sulfolobus tokodaii.
 CC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 CX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JCM 10545 / 7;
 RX PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Ankal A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermophilic
 RT crenarchaeon, Sulfolobus tokodaii strain7.";
 RL DNA Res. 8:123-140(2001).
 CC -1- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----

DR EMBL: AF000988; BAB67163.1; -
 DR PROSITE: PS00360; RIBOSOMAL_S9; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 137 AA; 15723 MW; 356170BF5A052C06 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 137;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 MEPLIA 43
 |||||
 Db 50 MEPLIA 56

RESULT 13

PER_DROPI STANDARD; PRT; 143 AA.
 AC 025206;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Period circadian protein (Fragment).
 GN PER.
 OS Drosophila picticornis (Fruit fly) (Idiomya picticornis).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_Taxid=7235;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95115533; PubMed-7815924;
 RA Nielsen J., Pelxoto A.A., Piccin A., Costa R., Kyriacou C.P.,
 RT Chalmers D.;
 RT "Big files, small repeats: the 'Thr-Gly' region of the period gene in
 RT Diptera.";
 RT Mol. Biol. Evol. 11:839-853(1994).
 CC -1- FUNCTION: ESSENTIAL FOR BIOLOGICAL CLOCK FUNCTIONS. DETERMINES THE
 CC PERIOD LENGTH OF CIRCADIAN AND ULTRADIAN RHYTHMS: AN INCREASE IN
 CC PER DOSAGE LEADS TO SHORTENED CIRCADIAN RHYTHMS AND A DECREASE IN
 CC LEADS TO LENGTHENED CIRCADIAN RHYTHMS. ESSENTIAL FOR THE CIRCADIAN
 CC RHYTHMICITY OF LOCOMOTOR ACTIVITY, ECLOSION BEHAVIOR, AND FOR THE
 CC RHYTHMIC COMPONENT OF THE MALE COURTSHIP SONG THAT ORIGINATES IN
 CC THE THORACIC NERVOUS SYSTEM. THE BIOLOGICAL CYCLE DEPENDS ON THE
 CC RHYTHMIC FORMATION AND NUCLEAR LOCALIZATION OF THE TIM-PER
 CC COMPLEX. LIGHT INDUCES THE DEGRADATION OF TIM, WHICH PROMOTES
 CC ELIMINATION OF PER. NUCLEAR ACTIVITY OF THE HETERODIMER
 CC COORDINATIVELY REGULATES PER AND TIM TRANSCRIPTION THROUGH A
 CC NEGATIVE FEEDBACK LOOP. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN
 CC TRANSCRIPTIONAL LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING
 CC INDIRECT TRANSCRIPTIONAL INHIBITION (BY SIMILARITY).
 CC -1- SUBUNIT: FORMS HETERODIMER WITH TIMELESS (TIM); THE COMPLEX THEN
 CC TRANSLLOCATES INTO THE NUCLEUS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEUS AT SPECIFIC PERIODS OF THE DAY.
 CC FIRST ACCUMULATES IN THE PERINUCLEAR REGION ABOUT ONE HOUR BEFORE
 CC TRANSLLOCATION INTO THE NUCLEUS. INTERACTION WITH TIM IS REQUIRED
 CC FOR NUCLEAR LOCALIZATION (BY SIMILARITY).
 CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
 CC DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
 CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
 CC PER-TIM (BY SIMILARITY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U11810; AAA76592.1; -
 DR FlyBase: Fggn0015136; DpicPer.
 KW Biological rhythms; Repeat; Nuclear protein; Phosphorylation.

FT NON_TER 1 1
 FT NON_TER 143 143
 SQ SEQUENCE 143 AA; 14459 MW; 5F4F2DC27F5D6846 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 143;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 AGPSLAA 29
 |||||
 Db 120 AGPSLAA 126

RESULT 14

YS72_MYCTU STANDARD; PRT; 147 AA.
 ID YS72_MYCTU
 AC 010800;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 16.6 kDa protein Rv2872.
 GN Rv2872 OR MT2939 OR MYCY274.03.
 GN Mycobacterium tuberculosis.
 OS Mycobacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Bacteria; Firmicutes; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NX NCBI_Taxid=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE-98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Sulton J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RT Nature 393:537-544(1998).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Eierschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Petersen J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z74024; CNA98372.1; -
 DR EMBL: AE007118; AAK47264.1; -
 DR TIGR: MT2939;
 DR TubercuList; Rv2872;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 147 AA; 16596 MW; F2DC73045F92553 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 147;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LAAPAM 33
 |||||
 Db 74 LAAPAM 80

RESULT 15
 YC52_PORPU

ID YC52_PORPU STANDARD; PRT; 174 AA.

AC P51192;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Hypothetical 20.1 kDa protein ycf52 (ORF174).

GN YCF52.

OS Porphyra purpurea.

OG Chloroplast.

OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.

OX NCBI_TaxID=2787;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AVONPORT;

RA Reith M.E., Munnoland J.;

RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast

genome."

RL Plant Mol. Biol. Rep. 13:33-335(1995).

CC -1- SIMILARITY: BELONGS TO THE ACETYLTRANSFERASE FAMILY. YCF52

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U38804; AAC08078.1; -

CC InterPro; IPR000182; Acetyltransf_GCN5.

CC Pfam; PF00583; Acetyltransf; 1.

CC Hypothetical protein; Transferase; Acyltransferase; Chloroplast.

CC SEQUENCE 174 AA; 2009 MW; 8F8F7E09301D87C2 CRC64;

CC -----

CC Query Match 1.3%; Score 7; DB 1; Length 174;

CC Best Local Similarity 100.0%; Pred. No. 21;

CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC QY 526 LIKOLRQ 532

CC |||||

CC Db 130 LIKOLRQ 136

CC Search completed: July 17, 2002, 16:48:35

CC Job time: 140 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 16:45:15 ; Search time 20.95 Seconds

(without alignments)
2463.006 Million cell updates/sec

Title: US-09-821-812-5

Perfect score: 537
Sequence: 1 GGLTTYIGIRLGVDRRLS.....PDHVDIOLIKOLRHRAML 537

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: - 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR.71:*

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	1.5	293	1 B9AG55	virB9 protein prec
2	8	1.5	532	2 D71267	probable UDP-N-ace
3	8	1.5	1332	1 I48314	homeotic protein C
4	7	1.3	59	2 AE1363	hypothetical prote
5	7	1.3	70	2 S56769	RNA-directed RNA p
6	7	1.3	75	1 R3EC18	ribosomal protein
7	7	1.3	75	1 E64076	ribosomal protein
8	7	1.3	75	2 B91276	30S ribosomal subu
9	7	1.3	75	2 E84995	30S ribosomal prot
10	7	1.3	75	2 B86117	30S ribosomal subu
11	7	1.3	75	2 AB0430	30S ribosomal prot
12	7	1.3	75	2 A11052	30S ribosomal chai
13	7	1.3	92	2 T49763	probable magnaport
14	7	1.3	110	2 S11786	nole protein - Rhl
15	7	1.3	116	2 B75456	hypothetical prote
16	7	1.3	116	2 C72654	hypothetical prote
17	7	1.3	133	2 D70105	hypothetical prote
18	7	1.3	140	2 S75420	probable ribosomal
19	7	1.3	147	2 C70923	hypothetical prote
20	7	1.3	147	2 G72640	hypothetical prote
21	7	1.3	151	2 G84131	hypothetical prote
22	7	1.3	174	2 S73113	hypothetical prote
23	7	1.3	183	2 T16186	hypothetical prote
24	7	1.3	209	2 E83270	hypothetical prote
25	7	1.3	220	2 C64146	hypothetical prote
26	7	1.3	221	2 JE0171	ribonuclease T2 (E
27	7	1.3	223	2 T37032	probable integral
28	7	1.3	229	2 I40068	probable ABC-type
29	7	1.3	234	2 AF0504	conserved hypotet

30	7	1.3	245	2 C84017	hypothetical prote
31	7	1.3	262	2 H75377	conserved hypotet
32	7	1.3	264	2 S28660	proliferating cell
33	7	1.3	267	2 AD1835	hypothetical prote
34	7	1.3	272	2 J01802	B8R 31K protein pr
35	7	1.3	272	2 G42526	B8R protein - vacc
36	7	1.3	274	2 T51714	probable formamido
37	7	1.3	281	2 S69525	hypothetical prote
38	7	1.3	282	2 D72679	hypothetical prote
39	7	1.3	283	2 AH3387	hypothetical cytos
40	7	1.3	284	2 C83237	probable transcrip
41	7	1.3	285	1 Y0EC88	fimbrial adhesin K
42	7	1.3	287	2 B87045	hypothetical prote
43	7	1.3	288	2 H70017	ribonuclease (EC 3
44	7	1.3	292	1 KFRB3	tissue factor prec
45	7	1.3	296	2 T45203	probable transport

ALIGNMENTS

RESULT 1
B9AG55
virB9 protein precursor - Agrobacterium tumefaciens plasmids pT15955 and pT1A6
C:Species: Agrobacterium tumefaciens
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
C:Accession: S00785; B30402; A27647; A35737; A26217
R:Thompson, D.V.; Melchers, L.S.; Idler, K.B.; Schlipperort, R.A.; Hooykaas, P.J.J.
Nucleic Acids Res. 16, 4621-4636, 1988
A:Title: Analysis of the complete nucleotide sequence of the Agrobacterium tumefaciens
A:Reference number: S00777; MUID:88247765
A:Accession: S00785
A:Molecule type: DNA
A:Residues: 1-293 <THO>
A:Cross-references: EMBL:X06826; NID:g39195; PID:CAA29979.1; PID:g757730
A:Experimental source: strain 15955, plasmid pT15955
R:Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.
submitted to GenBank, December 1989
A:Reference number: A30402
A:Accession: B30402
A:Molecule type: DNA
A:Residues: 1-293 <MAR>
A:Cross-references: GB:J03216; NID:q1196971; PID:AAA8654.1; PID:g1196981
A:Experimental source: plasmid pT1A6
R:Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.
J. Biol. Chem. 263, 5804-5814, 1988
A:Title: Characterization of the virB operon from an Agrobacterium tumefaciens T1 pla
A:Reference number: A28621; MUID:88186901
A:Accession: A27647
A:Molecule type: DNA
A:Residues: 1-271 <MA2>
A:Cross-references: GB:J03216
A:Experimental source: plasmid pT1A6
A:Note: this sequence was designated the amino-terminal portion of ORF 10 in this ref
J. Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.
J. Biol. Chem. 265, 4768, 1990
A:Reference number: A35737; MUID:90170994
A:Contents: erratum
A:Accession: A35737
A:Molecule type: DNA
A:Residues: 97-293 <MA3>
A:Experimental source: plasmid pT1A6
C:Genetics:
A:Genome: plasmid
C:Superfamily: tumor-inducing plasmid pTIC58 virB9 protein
F:1-21/Domain: signal sequence #status predicted <Sig>
F:22-293/Product: virB9 protein #status predicted <MAV>

Query Match 1.5%; Score 8; DB 1; Length 293;
Best local Similarity 100.0%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 LDOPVTL 300
Db 173 LDOPVTL 180

RESULT 2

D71267
probable UDP-N-acetylmuramylalanine-D-glutamate ligase (murD) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: D71267
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Ulterback, T.; McD
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: D71267
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-532 <COL>
A:Cross-references: GB:AE001259; GB:AE000520; NID:g3323209; PIDN:AAC65856.1; PID:g332321
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0903

Query Match 1.5%; Score 8; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 EOARALQ 477
Db 2 EOARALQ 9

RESULT 3

I48314
homeotic protein CDP - mouse
N:Alternate names: CCAAT displacement protein; homeotic protein Cux
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: I48314
R:Valarche, I.; Tissler-Seta, J.P.; Hirsch, M.R.; Martinez, S.; Goridis, C.; Brunet, J.F
Development 119, 881-896, 1993
A:Title: The mouse homeodomain protein Phox2 regulates Ncam promoter activity in concert
A:Reference number: I48314; MUID:94244481
A:Accession: I48314
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: mRNA
A:Residues: 1-1332 <RES>
A:Cross-references: EMBL:X75013; NID:g402589; PIDN:CAA52922.1; PID:g402590
C:Genetics:
A:Gene: Cux
C:Superfamily: homeotic protein CDP; cut repeat homology; homeobox homology
C:Keywords: DNA binding; duplication; homeobox; nucleus; transcription regulation
F:366-438/Domain: cut repeat homology <CU1>
F:755-827/Domain: cut repeat homology <CU2>
F:938-1010/Domain: cut repeat homology <CU3>
F:1057-1113/Domain: homeobox homology <HOX>

Query Match 1.5%; Score 8; DB 1; Length 1332;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 PSIAAPAA 32
Db 1230 PSIAAPAA 1237

RESULT 4

AE1363

hypothetical protein Imo2309 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AE1363

R:Glasier, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duhaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fahn,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkak, G.; Madueno, E.; Maltournam, A.;
ok, C.; Schluter, T.; Simes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1363
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-59 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00387.1; PID:g16411779; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: Imo2309

Query Match 1.3%; Score 7; DB 2; Length 59;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 AENEEA 418
Db 17 AENEEA 23

RESULT 5

S56769
RNA-directed RNA polymerase (EC 2.7.7.48) - Leishmania RNA virus 1-13 (fragment)
C:Species: Leishmania RNA virus 1-13
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 18-Jul-1997
C:Accession: S56769
R:Widmer, G.; Dooley, S.
Nucleic Acids Res. 23, 2300-2304, 1995
A:Title: Phylogenetic analysis of Leishmania RNA virus and Leishmania suggests anclen
A:Reference number: S56768; MUID:95334386
A:Accession: S56769
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-70 <WTD>
A:Cross-references: EMBL:L39069
C:Keywords: nucleotidyltransferase

Query Match 1.3%; Score 7; DB 2; Length 70;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 EIDVDE 94
Db 17 EIDVDE 23

RESULT 6

R3EC18
ribosomal protein S18 [validated] - Escherichia coli
C:Species: Escherichia coli
C:Date: 24-Apr-1984 #sequence_revision 24-May-1996 #text_change 28-Jan-2000
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from
A:Reference number: S56314; MUID:95334362
A:Accession: S56427
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-75 <BUR>
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97098.1; PID:g537043

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R.Yaguchi, M.
FEBS Lett. 59, 217-220, 1975
A>Title: Primary structure of protein S18 from the small Escherichia coli ribosomal subunit
A:Reference number: A02741; MUID:76210737
A:Accession: A02741
A:Molecule type: protein
A:Residues: 215, 'O', 17-75 <YAG>
A:Experimental source: strain K
R.Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E65231
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-75 <BLAT>
A:Cross-references: GB:AE000491; GB:U00096; NID:q2367357; PIDN:AMC77159.1; PID:q1790646;
A:Experimental source: strain K-12, substrain MG1655
R.Arnold, R.J.; Reilly, J.P.
Anal. Biochem. 269, 105-112, 1999
A>Title: Observation of Escherichia coli ribosomal proteins and their posttranslational
A:Reference number: A59071; MUID:99196679
A:Contents: annotation; mass spectrometric analysis
A>Note: mass spectrometric analysis of post-translational modifications; any acid labile
C:Comment: The amino end is acetylated by ribosomal-protein-alanine N-acetyltransferase
C:Genetics:
A:Gene: rpsR
A:Map position: 96 min
C:Complex: the ribosome is composed of the large (50S) and small (30S) subunit; the large
S RNA and 22 distinct proteins
C:Complex: small subunit ribosomal proteins: S1 (PIR:R3EC1), S2 (PIR:R3EC2), S3 (PIR:R3EC3), S4 (PIR:R3EC4), S5 (PIR:R3EC5), S6 (PIR:R3EC6), S7 (PIR:R3EC7), S8 (PIR:R3EC8), S9 (PIR:R3EC9), S10 (PIR:R3EC10), S11 (PIR:R3EC11), S12 (PIR:R3EC12), S13 (PIR:R3EC13), S14 (PIR:R3EC14), S15 (PIR:R3EC15), S16 (PIR:R3EC16), S17 (PIR:R3EC17), S18 (PIR:R3EC18), S19 (PIR:R3EC19), S20 (PIR:R3EC20), S21 (PIR:R3EC21), S22 (PIR:R3EC22), S23 (PIR:R3EC23), S24 (PIR:R3EC24), S25 (PIR:R3EC25), S26 (PIR:R3EC26), S27 (PIR:R3EC27), S28 (PIR:R3EC28), S29 (PIR:R3EC29), S30 (PIR:R3EC30), S31 (PIR:R3EC31), S32 (PIR:R3EC32), S33 (PIR:R3EC33), S34 (PIR:R3EC34), S35 (PIR:R3EC35), S36 (PIR:R3EC36), S37 (PIR:R3EC37), S38 (PIR:R3EC38), S39 (PIR:R3EC39), S40 (PIR:R3EC40), S41 (PIR:R3EC41), S42 (PIR:R3EC42), S43 (PIR:R3EC43), S44 (PIR:R3EC44), S45 (PIR:R3EC45), S46 (PIR:R3EC46), S47 (PIR:R3EC47), S48 (PIR:R3EC48), S49 (PIR:R3EC49), S50 (PIR:R3EC50), S51 (PIR:R3EC51), S52 (PIR:R3EC52), S53 (PIR:R3EC53), S54 (PIR:R3EC54), S55 (PIR:R3EC55), S56 (PIR:R3EC56), S57 (PIR:R3EC57), S58 (PIR:R3EC58), S59 (PIR:R3EC59), S60 (PIR:R3EC60), S61 (PIR:R3EC61), S62 (PIR:R3EC62), S63 (PIR:R3EC63), S64 (PIR:R3EC64), S65 (PIR:R3EC65), S66 (PIR:R3EC66), S67 (PIR:R3EC67), S68 (PIR:R3EC68), S69 (PIR:R3EC69), S70 (PIR:R3EC70), S71 (PIR:R3EC71), S72 (PIR:R3EC72), S73 (PIR:R3EC73), S74 (PIR:R3EC74), S75 (PIR:R3EC75), S76 (PIR:R3EC76), S77 (PIR:R3EC77), S78 (PIR:R3EC78), S79 (PIR:R3EC79), S80 (PIR:R3EC80), S81 (PIR:R3EC81), S82 (PIR:R3EC82), S83 (PIR:R3EC83), S84 (PIR:R3EC84), S85 (PIR:R3EC85), S86 (PIR:R3EC86), S87 (PIR:R3EC87), S88 (PIR:R3EC88), S89 (PIR:R3EC89), S90 (PIR:R3EC90), S91 (PIR:R3EC91), S92 (PIR:R3EC92), S93 (PIR:R3EC93), S94 (PIR:R3EC94), S95 (PIR:R3EC95), S96 (PIR:R3EC96), S97 (PIR:R3EC97), S98 (PIR:R3EC98), S99 (PIR:R3EC99), S100 (PIR:R3EC100), S101 (PIR:R3EC101), S102 (PIR:R3EC102), S103 (PIR:R3EC103), S104 (PIR:R3EC104), S105 (PIR:R3EC105), S106 (PIR:R3EC106), S107 (PIR:R3EC107), S108 (PIR:R3EC108), S109 (PIR:R3EC109), S110 (PIR:R3EC110), S111 (PIR:R3EC111), S112 (PIR:R3EC112), S113 (PIR:R3EC113), S114 (PIR:R3EC114), S115 (PIR:R3EC115), S116 (PIR:R3EC116), S117 (PIR:R3EC117), S118 (PIR:R3EC118), S119 (PIR:R3EC119), S120 (PIR:R3EC120), S121 (PIR:R3EC121), S122 (PIR:R3EC122), S123 (PIR:R3EC123), S124 (PIR:R3EC124), S125 (PIR:R3EC125), S126 (PIR:R3EC126), S127 (PIR:R3EC127), S128 (PIR:R3EC128), S129 (PIR:R3EC129), S130 (PIR:R3EC130), S131 (PIR:R3EC131), S132 (PIR:R3EC132), S133 (PIR:R3EC133), S134 (PIR:R3EC134), S135 (PIR:R3EC135), S136 (PIR:R3EC136), S137 (PIR:R3EC137), S138 (PIR:R3EC138), S139 (PIR:R3EC139), S140 (PIR:R3EC140), S141 (PIR:R3EC141), S142 (PIR:R3EC142), S143 (PIR:R3EC143), S144 (PIR:R3EC144), S145 (PIR:R3EC145), S146 (PIR:R3EC146), S147 (PIR:R3EC147), S148 (PIR:R3EC148), S149 (PIR:R3EC149), S150 (PIR:R3EC150), S151 (PIR:R3EC151), S152 (PIR:R3EC152), S153 (PIR:R3EC153), S154 (PIR:R3EC154), S155 (PIR:R3EC155), S156 (PIR:R3EC156), S157 (PIR:R3EC157), S158 (PIR:R3EC158), S159 (PIR:R3EC159), S160 (PIR:R3EC160), S161 (PIR:R3EC161), S162 (PIR:R3EC162), S163 (PIR:R3EC163), S164 (PIR:R3EC164), S165 (PIR:R3EC165), S166 (PIR:R3EC166), S167 (PIR:R3EC167), S168 (PIR:R3EC168), S169 (PIR:R3EC169), S170 (PIR:R3EC170), S171 (PIR:R3EC171), S172 (PIR:R3EC172), S173 (PIR:R3EC173), S174 (PIR:R3EC174), S175 (PIR:R3EC175), S176 (PIR:R3EC176), S177 (PIR:R3EC177), S178 (PIR:R3EC178), S179 (PIR:R3EC179), S180 (PIR:R3EC180), S181 (PIR:R3EC181), S182 (PIR:R3EC182), S183 (PIR:R3EC183), S184 (PIR:R3EC184), S185 (PIR:R3EC185), S186 (PIR:R3EC186), S187 (PIR:R3EC187), S188 (PIR:R3EC188), S189 (PIR:R3EC189), S190 (PIR:R3EC190), S191 (PIR:R3EC191), S192 (PIR:R3EC192), S193 (PIR:R3EC193), S194 (PIR:R3EC194), S195 (PIR:R3EC195), S196 (PIR:R3EC196), S197 (PIR:R3EC197), S198 (PIR:R3EC198), S199 (PIR:R3EC199), S200 (PIR:R3EC200), S201 (PIR:R3EC201), S202 (PIR:R3EC202), S203 (PIR:R3EC203), S204 (PIR:R3EC204), S205 (PIR:R3EC205), S206 (PIR:R3EC206), S207 (PIR:R3EC207), S208 (PIR:R3EC208), S209 (PIR:R3EC209), S210 (PIR:R3EC210), S211 (PIR:R3EC211), S212 (PIR:R3EC212), S213 (PIR:R3EC213), S214 (PIR:R3EC214), S215 (PIR:R3EC215), S216 (PIR:R3EC216), S217 (PIR:R3EC217), S218 (PIR:R3EC218), S219 (PIR:R3EC219), S220 (PIR:R3EC220), S221 (PIR:R3EC221), S222 (PIR:R3EC222), S223 (PIR:R3EC223), S224 (PIR:R3EC224), S225 (PIR:R3EC225), S226 (PIR:R3EC226), S227 (PIR:R3EC227), S228 (PIR:R3EC228), S229 (PIR:R3EC229), S230 (PIR:R3EC230), S231 (PIR:R3EC231), S232 (PIR:R3EC232), S233 (PIR:R3EC233), S234 (PIR:R3EC234), S235 (PIR:R3EC235), S236 (PIR:R3EC236), S237 (PIR:R3EC237), S238 (PIR:R3EC238), S239 (PIR:R3EC239), S240 (PIR:R3EC240), S241 (PIR:R3EC241), S242 (PIR:R3EC242), S243 (PIR:R3EC243), S244 (PIR:R3EC244), S245 (PIR:R3EC245), S246 (PIR:R3EC246), S247 (PIR:R3EC247), S248 (PIR:R3EC248), S249 (PIR:R3EC249), S250 (PIR:R3EC250), S251 (PIR:R3EC251), S252 (PIR:R3EC252), S253 (PIR:R3EC253), S254 (PIR:R3EC254), S255 (PIR:R3EC255), S256 (PIR:R3EC256), S257 (PIR:R3EC257), S258 (PIR:R3EC258), S259 (PIR:R3EC259), S260 (PIR:R3EC260), S261 (PIR:R3EC261), S262 (PIR:R3EC262), S263 (PIR:R3EC263), S264 (PIR:R3EC264), S265 (PIR:R3EC265), S266 (PIR:R3EC266), S267 (PIR:R3EC267), S268 (PIR:R3EC268), S269 (PIR:R3EC269), S270 (PIR:R3EC270), S271 (PIR:R3EC271), S272 (PIR:R3EC272), S273 (PIR:R3EC273), S274 (PIR:R3EC274), S275 (PIR:R3EC275), S276 (PIR:R3EC276), S277 (PIR:R3EC277), S278 (PIR:R3EC278), S279 (PIR:R3EC279), S280 (PIR:R3EC280), S281 (PIR:R3EC281), S282 (PIR:R3EC282), S283 (PIR:R3EC283), S284 (PIR:R3EC284), S285 (PIR:R3EC285), S286 (PIR:R3EC286), S287 (PIR:R3EC287), S288 (PIR:R3EC288), S289 (PIR:R3EC289), S290 (PIR:R3EC290), S291 (PIR:R3EC291), S292 (PIR:R3EC292), S293 (PIR:R3EC293), S294 (PIR:R3EC294), S295 (PIR:R3EC295), S296 (PIR:R3EC296), S297 (PIR:R3EC297), S298 (PIR:R3EC298), S299 (PIR:R3EC299), S300 (PIR:R3EC300), S301 (PIR:R3EC301), S302 (PIR:R3EC302), S303 (PIR:R3EC303), S304 (PIR:R3EC304), S305 (PIR:R3EC305), S306 (PIR:R3EC306), S307 (PIR:R3EC307), S308 (PIR:R3EC308), S309 (PIR:R3EC309), S310 (PIR:R3EC310), S311 (PIR:R3EC311), S312 (PIR:R3EC312), S313 (PIR:R3EC313), S314 (PIR:R3EC314), S315 (PIR:R3EC315), S316 (PIR:R3EC316), S317 (PIR:R3EC317), S318 (PIR:R3EC318), S319 (PIR:R3EC319), S320 (PIR:R3EC320), S321 (PIR:R3EC321), S322 (PIR:R3EC322), S323 (PIR:R3EC323), S324 (PIR:R3EC324), S325 (PIR:R3EC325), S326 (PIR:R3EC326), S327 (PIR:R3EC327), S328 (PIR:R3EC328), S329 (PIR:R3EC329), S330 (PIR:R3EC330), S331 (PIR:R3EC331), S332 (PIR:R3EC332), S333 (PIR:R3EC333), S334 (PIR:R3EC334), S335 (PIR:R3EC335), S336 (PIR:R3EC336), S337 (PIR:R3EC337), S338 (PIR:R3EC338), S339 (PIR:R3EC339), S340 (PIR:R3EC340), S341 (PIR:R3EC341), S342 (PIR:R3EC342), S343 (PIR:R3EC343), S344 (PIR:R3EC344), S345 (PIR:R3EC345), S346 (PIR:R3EC346), S347 (PIR:R3EC347), S348 (PIR:R3EC348), S349 (PIR:R3EC349), S350 (PIR:R3EC350), S351 (PIR:R3EC351), S352 (PIR:R3EC352), S353 (PIR:R3EC353), S354 (PIR:R3EC354), S355 (PIR:R3EC355), S356 (PIR:R3EC356), S357 (PIR:R3EC357), S358 (PIR:R3EC358), S359 (PIR:R3EC359), S360 (PIR:R3EC360), S361 (PIR:R3EC361), S362 (PIR:R3EC362), S363 (PIR:R3EC363), S364 (PIR:R3EC364), S365 (PIR:R3EC365), S366 (PIR:R3EC366), S367 (PIR:R3EC367), S368 (PIR:R3EC368), S369 (PIR:R3EC369), S370 (PIR:R3EC370), S371 (PIR:R3EC371), S372 (PIR:R3EC372), S373 (PIR:R3EC373), S374 (PIR:R3EC374), S375 (PIR:R3EC375), S376 (PIR:R3EC376), S377 (PIR:R3EC377), S378 (PIR:R3EC378), S379 (PIR:R3EC379), S380 (PIR:R3EC380), S381 (PIR:R3EC381), S382 (PIR:R3EC382), S383 (PIR:R3EC383), S384 (PIR:R3EC384), S385 (PIR:R3EC385), S386 (PIR:R3EC386), S387 (PIR:R3EC387), S388 (PIR:R3EC388), S389 (PIR:R3EC389), S390 (PIR:R3EC390), S391 (PIR:R3EC391), S392 (PIR:R3EC392), S393 (PIR:R3EC393), S394 (PIR:R3EC394), S395 (PIR:R3EC395), S396 (PIR:R3EC396), S397 (PIR:R3EC397), S398 (PIR:R3EC398), S399 (PIR:R3EC399), S400 (PIR:R3EC400), S401 (PIR:R3EC401), S402 (PIR:R3EC402), S403 (PIR:R3EC403), S404 (PIR:R3EC404), S405 (PIR:R3EC405), S406 (PIR:R3EC406), S407 (PIR:R3EC407), S408 (PIR:R3EC408), S409 (PIR:R3EC409), S410 (PIR:R3EC410), S411 (PIR:R3EC411), S412 (PIR:R3EC412), S413 (PIR:R3EC413), S414 (PIR:R3EC414), S415 (PIR:R3EC415), S416 (PIR:R3EC416), S417 (PIR:R3EC417), S418 (PIR:R3EC418), S419 (PIR:R3EC419), S420 (PIR:R3EC420), S421 (PIR:R3EC421), S422 (PIR:R3EC422), S423 (PIR:R3EC423), S424 (PIR:R3EC424), S425 (PIR:R3EC425), S426 (PIR:R3EC426), S427 (PIR:R3EC427), S428 (PIR:R3EC428), S429 (PIR:R3EC429), S430 (PIR:R3EC430), S431 (PIR:R3EC431), S432 (PIR:R3EC432), S433 (PIR:R3EC433), S434 (PIR:R3EC434), S435 (PIR:R3EC435), S436 (PIR:R3EC436), S437 (PIR:R3EC437), S438 (PIR:R3EC438), S439 (PIR:R3EC439), S440 (PIR:R3EC440), S441 (PIR:R3EC441), S442 (PIR:R3EC442), S443 (PIR:R3EC443), S444 (PIR:R3EC444), S445 (PIR:R3EC445), S446 (PIR:R3EC446), S447 (PIR:R3EC447), S448 (PIR:R3EC448), S449 (PIR:R3EC449), S450 (PIR:R3EC450), S451 (PIR:R3EC451), S452 (PIR:R3EC452), S453 (PIR:R3EC453), S454 (PIR:R3EC454), S455 (PIR:R3EC455), S456 (PIR:R3EC456), S457 (PIR:R3EC457), S458 (PIR:R3EC458), S459 (PIR:R3EC459), S460 (PIR:R3EC460), S461 (PIR:R3EC461), S462 (PIR:R3EC462), S463 (PIR:R3EC463), S464 (PIR:R3EC464), S465 (PIR:R3EC465), S466 (PIR:R3EC466), S467 (PIR:R3EC467), S468 (PIR:R3EC468), S469 (PIR:R3EC469), S470 (PIR:R3EC470), S471 (PIR:R3EC471), S472 (PIR:R3EC472), S473 (PIR:R3EC473), S474 (PIR:R3EC474), S475 (PIR:R3EC475), S476 (PIR:R3EC476), S477 (PIR:R3EC477), S478 (PIR:R3EC478), S479 (PIR:R3EC479), S480 (PIR:R3EC480), S481 (PIR:R3EC481), S482 (PIR:R3EC482), S483 (PIR:R3EC483), S484 (PIR:R3EC484), S485 (PIR:R3EC485), S486 (PIR:R3EC486), S487 (PIR:R3EC487), S488 (PIR:R3EC488), S489 (PIR:R3EC489), S490 (PIR:R3EC490), S491 (PIR:R3EC491), S492 (PIR:R3EC492), S493 (PIR:R3EC493), S494 (PIR:R3EC494), S495 (PIR:R3EC495), S496 (PIR:R3EC496), S497 (PIR:R3EC497), S498 (PIR:R3EC498), S499 (PIR:R3EC499), S500 (PIR:R3EC500), S501 (PIR:R3EC501), S502 (PIR:R3EC502), S503 (PIR:R3EC503), S504 (PIR:R3EC504), S505 (PIR:R3EC505), S506 (PIR:R3EC506), S507 (PIR:R3EC507), S508 (PIR:R3EC508), S509 (PIR:R3EC509), S510 (PIR:R3EC510), S511 (PIR:R3EC511), S512 (PIR:R3EC512), S513 (PIR:R3EC513), S514 (PIR:R3EC514), S515 (PIR:R3EC515), S516 (PIR:R3EC516), S517 (PIR:R3EC517), S518 (PIR:R3EC518), S519 (PIR:R3EC519), S520 (PIR:R3EC520), S521 (PIR:R3EC521), S522 (PIR:R3EC522), S523 (PIR:R3EC523), S524 (PIR:R3EC524), S525 (PIR:R3EC525), S526 (PIR:R3EC526), S527 (PIR:R3EC527), S528 (PIR:R3EC528), S529 (PIR:R3EC529), S530 (PIR:R3EC530), S531 (PIR:R3EC531), S532 (PIR:R3EC532), S533 (PIR:R3EC533), S534 (PIR:R3EC534), S535 (PIR:R3EC535), S536 (PIR:R3EC536), S537 (PIR:R3EC537), S538 (PIR:R3EC538), S539 (PIR:R3EC539), S540 (PIR:R3EC540), S541 (PIR:R3EC541), S542 (PIR:R3EC542), S543 (PIR:R3EC543), S544 (PIR:R3EC544), S545 (PIR:R3EC545), S546 (PIR:R3EC546), S547 (PIR:R3EC547), S548 (PIR:R3EC548), S549 (PIR:R3EC549), S550 (PIR:R3EC550), S551 (PIR:R3EC551), S552 (PIR:R3EC552), S553 (PIR:R3EC553), S554 (PIR:R3EC554), S555 (PIR:R3EC555), S556 (PIR:R3EC556), S557 (PIR:R3EC557), S558 (PIR:R3EC558), S559 (PIR:R3EC559), S560 (PIR:R3EC560), S561 (PIR:R3EC561), S562 (PIR:R3EC562), S563 (PIR:R3EC563), S564 (PIR:R3EC564), S565 (PIR:R3EC565), S566 (PIR:R3EC566), S567 (PIR:R3EC567), S568 (PIR:R3EC568), S569 (PIR:R3EC569), S570 (PIR:R3EC570), S571 (PIR:R3EC571), S572 (PIR:R3EC572), S573 (PIR:R3EC573), S574 (PIR:R3EC574), S575 (PIR:R3EC575), S576 (PIR:R3EC576), S577 (PIR:R3EC577), S578 (PIR:R3EC578), S579 (PIR:R3EC579), S580 (PIR:R3EC580), S581 (PIR:R3EC581), S582 (PIR:R3EC582), S583 (PIR:R3EC583), S584 (PIR:R3EC584), S585 (PIR:R3EC585), S586 (PIR:R3EC586), S587 (PIR:R3EC587), S588 (PIR:R3EC588), S589 (PIR:R3EC589), S590 (PIR:R3EC590), S591 (PIR:R3EC591), S592 (PIR:R3EC592), S593 (PIR:R3EC593), S594 (PIR:R3EC594), S595 (PIR:R3EC595), S596 (PIR:R3EC596), S597 (PIR:R3EC597), S598 (PIR:R3EC598), S599 (PIR:R3EC599), S600 (PIR:R3EC600), S601 (PIR:R3EC601), S602 (PIR:R3EC602), S603 (PIR:R3EC603), S604 (PIR:R3EC604), S605 (PIR:R3EC605), S606 (PIR:R3EC606), S607 (PIR:R3EC607), S608 (PIR:R3EC608), S609 (PIR:R3EC609), S610 (PIR:R3EC610), S611 (PIR:R3EC611), S612 (PIR:R3EC612), S613 (PIR:R3EC613), S614 (PIR:R3EC614), S615 (PIR:R3EC615), S616 (PIR:R3EC616), S617 (PIR:R3EC617), S618 (PIR:R3EC618), S619 (PIR:R3EC619), S620 (PIR:R3EC620), S621 (PIR:R3EC621), S622 (PIR:R3EC622), S623 (PIR:R3EC623), S624 (PIR:R3EC624), S625 (PIR:R3EC625), S626 (PIR:R3EC626), S627 (PIR:R3EC627), S628 (PIR:R3EC628), S629 (PIR:R3EC629), S630 (PIR:R3EC630), S631 (PIR:R3EC631), S632 (PIR:R3EC632), S633 (PIR:R3EC633), S634 (PIR:R3EC634), S635 (PIR:R3EC635), S636 (PIR:R3EC636), S637 (PIR:R3EC637), S638 (PIR:R3EC638), S639 (PIR:R3EC639), S640 (PIR:R3EC640), S641 (PIR:R3EC641), S642 (PIR:R3EC642), S643 (PIR:R3EC643), S644 (PIR:R3EC644), S645 (PIR:R3EC645), S646 (PIR:R3EC646), S647 (PIR:R3EC647), S648 (PIR:R3EC648), S649 (PIR:R3EC649), S650 (PIR:R3EC650), S651 (PIR:R3EC651), S652 (PIR:R3EC652), S653 (PIR:R3EC653), S654 (PIR:R3EC654), S655 (PIR:R3EC655), S656 (PIR:R3EC656), S657 (PIR:R3EC657), S658 (PIR:R3EC658), S659 (PIR:R3EC659), S660 (PIR:R3EC660), S661 (PIR:R3EC661), S662 (PIR:R3EC662), S663 (PIR:R3EC663), S664 (PIR:R3EC664), S665 (PIR:R3EC665), S666 (PIR:R3EC666), S667 (PIR:R3EC667), S668 (PIR:R3EC668), S669 (PIR:R3EC669), S670 (PIR:R3EC670), S671 (PIR:R3EC671), S672 (PIR:R3EC672), S673 (PIR:R3EC673), S674 (PIR:R3EC674), S675 (PIR:R3EC675), S676 (PIR:R3EC676), S677 (PIR:R3EC677), S678 (PIR:R3EC678), S679 (PIR:R3EC679), S680 (PIR:R3EC680), S681 (PIR:R3EC681), S682 (PIR:R3EC682), S683 (PIR:R3EC683), S684 (PIR:R3EC684), S685 (PIR:R3EC685), S686 (PIR:R3EC686), S687 (PIR:R3EC687), S688 (PIR:R3EC688), S689 (PIR:R3EC689), S690 (PIR:R3EC690), S691 (PIR:R3EC691), S692 (PIR:R3EC692), S693 (PIR:R3EC693), S694 (PIR:R3EC694), S695 (PIR:R3EC695), S696 (PIR:R3EC696), S697 (PIR:R3EC697), S698 (PIR:R3EC698), S699 (PIR:R3EC699), S700 (PIR:R3EC700), S701 (PIR:R3EC701), S702 (PIR:R3EC702), S703 (PIR:R3EC703), S704 (PIR:R3EC704), S705 (PIR:R3EC705), S706 (PIR:R3EC706), S707 (PIR:R3EC707), S708 (PIR:R3EC708), S709 (PIR:R3EC709), S710 (PIR:R3EC710), S711 (PIR:R3EC711), S712 (PIR:R3EC712), S713 (PIR:R3EC713), S714 (PIR:R3EC714), S715 (PIR:R3EC715), S716 (PIR:R3EC716), S717 (PIR:R3EC717), S718 (PIR:R3EC718), S719 (PIR:R3EC719), S720 (PIR:R3EC720), S721 (PIR:R3EC721), S722 (PIR:R3EC722), S723 (PIR:R3EC723), S724 (PIR:R3EC724), S725 (PIR:R3EC725), S726 (PIR:R3EC726), S727 (PIR:R3EC727), S728 (PIR:R3EC728), S729 (PIR:R3EC729), S730 (PIR:R3EC730), S731 (PIR:R3EC731), S732 (PIR:R3EC732), S733 (PIR:R3EC733), S734 (PIR:R3EC734), S735 (PIR:R3EC735), S736 (PIR:R3EC736), S737 (PIR:R3EC737), S738 (PIR:R3EC738), S739 (PIR:R3EC739), S740 (PIR:R3EC740), S741 (PIR:R3EC741), S742 (PIR:R3EC742), S743 (PIR:R3EC743), S744 (PIR:R3EC744), S745 (PIR:R3EC745), S746 (PIR:R3EC746), S747 (PIR:R3EC747), S748 (PIR:R3EC748), S749 (PIR:R3EC749), S750 (PIR:R3EC750), S751 (PIR:R3EC751), S752 (PIR:R3EC752), S753 (PIR:R3EC753), S754 (PIR:R3EC754), S755 (PIR:R3EC755), S756 (PIR:R3EC756), S757 (PIR:R3EC757), S758 (PIR:R3EC758), S759 (PIR:R3EC759), S760 (PIR:R3EC760), S761 (PIR:R3EC761), S762 (PIR:R3EC762), S763 (PIR:R3EC763), S764 (PIR:R3EC764), S765 (PIR:R3EC765), S766 (PIR:R3EC766), S767 (PIR:R3EC767), S768 (PIR:R3EC768), S769 (PIR:R3EC769), S770 (PIR:R3EC770), S771 (PIR:R3EC771), S772 (PIR:R3EC772), S773 (PIR:R3EC773), S774 (PIR:R3EC774), S775 (PIR:R3EC775), S776 (PIR:R3EC776), S777 (PIR:R3EC777), S778 (PIR:R3EC778), S779 (PIR:R3EC779), S780 (PIR:R3EC780), S781 (PIR:R3EC781), S782 (PIR:R3EC782), S783 (PIR:R3EC783), S784 (PIR:R3EC784), S785 (PIR:R3EC785), S786 (PIR:R3EC7

C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: B86117
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: AB5480; MUID:21074935; PMID:11206551
 A:Accession: B86117
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-75 <STD>
 A:Cross-references: GB:AE005174; NID:g12519193; PIDN:AAG59398.1; GSPDB:GN00145; UWGP:Z58
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: rpsR

Query Match 1.3%; Score 7; DB 2; Length 75;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 YFRRKKF 52
 |||||
 Db 4 YFRRKKF 10

RESULT 11

AB0430
 30S ribosomal protein S18 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C:Accession: AB0430
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Terraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AB0430
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-75 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC92766.1; PID:g15981461; GSPDB:GN00175
 C:Genetics:
 A:Gene: rpsR

Query Match 1.3%; Score 7; DB 2; Length 75;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 YFRRKKF 52
 |||||
 Db 4 YFRRKKF 10

RESULT 12

AI1052

30s ribosomal chain protein S18 [imported] - Salmonella enterica subsp. enterica serovar
 C:Species: Salmonella enterica subsp. enterica serovar Typhl
 A:Note: this species has also been called Salmonella typhl
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AI1052
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AI1052
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-75 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD06870.1; PID:g16505518; GSPDB:GN00176
 C:Genetics:
 A:Gene: rpsR

Query Match 1.3%; Score 7; DB 2; Length 75;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 YFRRKKF 52
 |||||
 Db 4 YFRRKKF 10

RESULT 13

T49763
 probable magnaporin protein [imported] - Neurospora crassa

N:Alternate names: protein B24B19.350
 C:Species: Neurospora crassa
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 03-Nov-2000
 C:Accession: T49763
 R:Schulte, U.; Align, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
 submitted to the Protein Sequence Database, May 2000
 A:Reference number: Z25022
 A:Accession: T49763
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-92 <SCH>
 A:Cross-references: EMBL:AL356192; GSPDB:GN00116; NCSP:B24B19.350
 A:Experimental source: BAC clone B24B19; strain OR74A
 C:Genetics:
 A:Gene: NCSP:B24B19.350
 A:Map position: 6
 A:Insertions: 56/1
 C:Superfamily: hydrophobin HPBI

Query Match 1.3%; Score 7; DB 2; Length 92;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 LAAPDAM 33
 |||||
 Db 16 LAAPDAM 22

RESULT 14

S11786
 nole protein - Rhizobium leguminosarum bv. phaseol

C:Species: Rhizobium leguminosarum bv. phaseol
 C:Date: 03-May-1994 #sequence_revision 20-Feb-1995 #text_change 21-Jul-2000
 C:Accession: S11786
 R:Davis, E.O.; Johnston, A.W.B.
 Mol. Microbiol. 4, 921-932, 1990
 A:Title: Analysis of three nod genes in Rhizobium leguminosarum biovar phaseol.
 A:Reference number: S11786; MUID:91014692
 A:Accession: S11786
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-110 <DAV>
 A:Cross-references: GB:X54214; NID:g46231; PIDN:CA38125.1; PID:g46232
 C:Genetics:
 A:Gene: nole
 C:Superfamily: Rhizobium nole protein

Query Match 1.3%; Score 7; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 AGPSIAA 29

DB 20 AGPSLAA 26

|||||||
RESULT 15

B75456 hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000

C:Accession: B75456

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Yamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; M01D:20036896

A:Accession: B75456

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-116 <WHI>

A:Cross-references: GB:AE001947; GB:AE000513; NID:g6458665; PIDN:AAF10529.1; PID:g645867

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0947

A:Map position: 1

C:Superfamily: Deinococcus radiodurans hypothetical protein DR0947

Query Match 1.3%; Score 7; DB 2; Length 116;
Best Local Similarity 100.0%; Pred.No. 40;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 468 HVEQARA 474
DB 8 HVEQARA 14

Search completed: July 17, 2002, 16:47:38
Job time: 143 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 16:44:45 ; Search time 15.3 Seconds
(without alignments)
857.290 Million cell updates/sec

Title: US-09-821-812-5

Perfect score: 537

Sequence: 1 GGLTTVIGTFLGVDRPRLS.....PDHDTOLIKOLRQHFAML 537

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	1.3	29	1	US-08-363-475-1
2	7	1.3	215	3	US-08-462-778-2
3	7	1.3	267	4	US-08-818-112-142
4	7	1.3	267	4	US-08-818-111-137
5	7	1.3	267	4	US-09-056-556-142
6	7	1.3	315	3	US-09-135-639-4
7	7	1.3	437	2	US-08-883-515-4
8	7	1.3	498	1	US-08-457-274A-24
9	7	1.3	498	5	PCT-US95-05758-24
10	7	1.3	507	1	US-08-363-475-19
11	7	1.3	532	1	US-08-363-475-22
12	7	1.3	565	2	US-08-700-548-2
13	7	1.3	566	2	US-08-335-8659-9
14	7	1.3	593	3	US-08-335-8659-21
15	7	1.3	1112	4	US-09-353-585-2
16	7	1.3	1112	4	US-09-353-585-3
17	7	1.3	2938	5	PCT-US94-00198-3
18	6	1.1	9	2	US-08-146-028-342
19	6	1.1	9	2	US-08-146-028-343
20	6	1.1	9	2	US-08-146-028-344
21	6	1.1	9	2	US-08-146-028-345
22	6	1.1	9	2	US-08-146-028-348
23	6	1.1	9	2	US-08-146-028-349
24	6	1.1	9	2	US-08-146-028-350
25	6	1.1	9	2	US-08-146-028-351
26	6	1.1	9	4	US-08-723-425A-342
27	6	1.1	9	4	US-08-723-425A-343

28	6	1.1	9	4	US-08-723-425A-344	Sequence 344, App
29	6	1.1	9	4	US-08-723-425A-345	Sequence 345, App
30	6	1.1	9	4	US-08-723-425A-348	Sequence 348, App
31	6	1.1	9	4	US-08-723-425A-349	Sequence 349, App
32	6	1.1	9	4	US-08-723-425A-350	Sequence 350, App
33	6	1.1	9	4	US-08-723-425A-351	Sequence 351, App
34	6	1.1	9	4	US-09-112-206-342	Sequence 342, App
35	6	1.1	9	4	US-09-112-206-343	Sequence 343, App
36	6	1.1	9	4	US-09-112-206-344	Sequence 344, App
37	6	1.1	9	4	US-09-112-206-345	Sequence 345, App
38	6	1.1	9	4	US-09-112-206-348	Sequence 348, App
39	6	1.1	9	4	US-09-112-206-349	Sequence 349, App
40	6	1.1	9	4	US-09-112-206-350	Sequence 350, App
41	6	1.1	9	4	US-09-112-206-351	Sequence 351, App
42	6	1.1	10	3	US-08-159-339A-461	Sequence 461, App
43	6	1.1	20	2	US-08-466-975A-14	Sequence 14, App1
44	6	1.1	20	2	US-08-466-975A-15	Sequence 15, App1
45	6	1.1	20	2	US-08-391-671A-14	Sequence 14, App1

ALIGNMENTS

RESULT 1
US-08-363-475-1
Sequence 1, Application US/08363475
Patent No. 5516679
GENERAL INFORMATION:
APPLICANT: Chlang, Shu-Jen
APPLICANT: Burnett Jr., William V.
TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM
TITLE OF INVENTION: FUSARIUM OXYSPORIUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Thomas R. Savitsky
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/363,475
FILING DATE: 23-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Savitsky, Thomas R.
REGISTRATION NUMBER: 31,661
REFERENCE/DOCKET NUMBER: ON-0134
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252 4956
TELEFAX: (609) 252 4526
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: Linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Fusarium oxysporum
FEATURE:
NAME/KEY: Peptide
LOCATION: 7
OTHER INFORMATION: /note= "Amino acid 7 can be ala or
OTHER INFORMATION: Lys"

NAME/KEY: Peptide
LOCATION: 22
OTHER INFORMATION: /note="Amino acid 22 can be Thr or
OTHER INFORMATION: Val"
US-08-363-475-1

Query Match 1.3%; Score 7; DB 1; Length 29;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 115 SLKUPGT 121
DB 15 SLKUPGT 21

RESULT 2
US-08-462-778-2
Sequence 2, Application US/08462778
Patent No. 6077517
GENERAL INFORMATION:
APPLICANT: Thomas, Wayne R.
APPLICANT: Chua, Kaw-Yan
TITLE OF INVENTION: Allergenic Protein and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25.
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,778
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/031,141
FILING DATE: 12 March 1993
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 215 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-462-778-2

Query Match 1.3%; Score 7; DB 3; Length 215;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 400 MTLTSEE 406
DB 152 MTLTSEE 158

RESULT 3
US-08-818-112-142
Sequence 142, Application US/08818112

Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
TITLE OF INVENTION: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 267 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-818-112-142

Query Match 1.3%; Score 7; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 VPRPGTS 115
DB 78 VPRPGTS 84

RESULT 4
US-08-818-111-137
Sequence 137, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TITLE OF INVENTION: TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 137:
SEQUENCE CHARACTERISTICS:
LENGTH: 267 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-818-111-137

Query Match 1.3%; Score 7; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 VPRPGTS 115
Db 78 VPRPGTS 84

RESULT 5
US-09-056-556-142
Sequence 142, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 267 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-056-556-142

Query Match 1.3%; Score 7; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 VPRPGTS 115
Db 78 VPRPGTS 84

RESULT 6
US-09-135-639-4
Sequence 4, Application US/09135639
Patent No. 6004793
GENERAL INFORMATION:
APPLICANT: LUNNEN, KEITH D.
APPLICANT: DALTON MICHAEL A.,
APPLICANT: WILSON, GEOFFREY G.
APPLICANT: XU, SHUANG-YONG
TITLE OF INVENTION: Method for Cloning and Producing The Avai Restriction
TITLE OF INVENTION: Recombinant Avai Restriction Endonuclease
FILE REFERENCE: Avai
CURRENT APPLICATION NUMBER: US/09/135,639
CURRENT FILING DATE: 1998-08-18
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 4
LENGTH: 315
TYPE: PRT
ORGANISM: Anabaena variabilis
US-09-135-639-4

Query Match 1.3%; Score 7; DB 3; Length 315;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VEQARAL 475
Db 39 VEQARAL 45

RESULT 7
US-08-883-515-4
Sequence 4, Application US/08883515
Patent No. 5981836
GENERAL INFORMATION:
APPLICANT: Ostryoung, Katherine W
TITLE OF INVENTION: PLANT CHLOROPLAST DIVISION GENES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Plinkney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/883,515
FILING DATE:
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 920905.90016
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9186
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 437 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-883-515-4

Query Match
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 508 VAAQKSE 514
|||||
DB 7 VAAQKSE 13

RESULT 8
US-08-457-274A-24
Sequence 24, Application US/08457274A
Patent No. 5734086
GENERAL INFORMATION:
APPLICANT: Scott, Jeffrey G.
APPLICANT: Tomita, Takashi
TITLE OF INVENTION: Cytochrome P4501pr Gene and Its Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,274A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/240 (D-1519)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1304
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 498 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: papillio polyxnes
STRAIN:
DEVELOPMENTAL STAGE: Adult
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
US-08-457-274A-24

Query Match
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 461 VLEMRKG 467
|||||
DB 245 VLEMRKG 251

RESULT 9
PCT-US95-05758-24
Sequence 24, Application PC/TUS9505758
GENERAL INFORMATION:
APPLICANT: Cornell Research Foundation, Inc.
TITLE OF INVENTION: Cytochrome P4501pr Gene and Its
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05758
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/241 (D-1519)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1304
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 498 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: papillio polyxnes
STRAIN:
DEVELOPMENTAL STAGE: Adult
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
PCT-US95-05758-24

Query Match
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 461 VLEMRKG 467
|||||
DB 245 VLEMRKG 251

RESULT 10
US-08-363-475-19
Sequence 19, Application US/08363475
Patent No. 5516679

GENERAL INFORMATION:
APPLICANT: Chiang, Shu-Jen
APPLICANT: Burnett Jr., William V.
APPLICANT: Tonzi, Sean M.
TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM
TITLE OF INVENTION: FUSARIUM OXYSPORUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas R. Savitsky
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/363,475
FILING DATE: 23-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Savitsky, Thomas R.
REGISTRATION NUMBER: 31,661
REFERENCE/DOCKET NUMBER: ON-0134
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252 4956
TELEFAX: (609) 252 4526
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-363-475-19

Query Match 1.3%; Score 7; DB 1; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SLKLPGT 121
DB 15 SLKLPGT 21

RESULT 11
US-08-363-475-22
Sequence 22, Application US/08363475
Patent No. 5516679
GENERAL INFORMATION:
APPLICANT: Chiang, Shu-Jen
APPLICANT: Burnett Jr., William V.
APPLICANT: Tonzi, Sean M.
TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM
TITLE OF INVENTION: FUSARIUM OXYSPORUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas R. Savitsky
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/363,475

FILING DATE: 23-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Savitsky, Thomas R.
REGISTRATION NUMBER: 31,661
REFERENCE/DOCKET NUMBER: ON-0134
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252 4956
TELEFAX: (609) 252 4526
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-363-475-22

Query Match 1.3%; Score 7; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SLKLPGT 121
DB 40 SLKLPGT 46

RESULT 12
US-08-700-548-2
Sequence 2, Application US/08700548
Patent No. 5910310
GENERAL INFORMATION:
APPLICANT: Heinen, Ernst; Schmeer, No. 5910310bert; Herbst, Werner
TITLE OF INVENTION: Para-Influenza Virus-Containing Vaccines for
TITLE OF INVENTION: Preventing Porcine Reproductive and Respiratory
TITLE OF INVENTION: Syndrome
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh 6500
OPERATING SYSTEM: System 7.5
SOFTWARE: Wordperfect 3.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/700,548
FILING DATE: 30-AUG-1996
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: PCT/EP95/00642
FILING DATE: 22-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 44 07 489.1
FILING DATE: 07-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: BAYER 9673-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-700-548-2

Query Match 1.3%; Score 7; DB 2; Length 565;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 QVPRPGT 114
1111111
Db 439 QVPRPGT 445

RESULT 13
US-08-335-865J-9
; Sequence 9, Application US/08335865J
; Patent No. 6107472
; GENERAL INFORMATION:
; APPLICANT: Stackner, Steven A.; Hovens, Christopher M.,
; APPLICANT: Wilks, Andrew F.
; TITLE OF INVENTION: RECEPTOR-TYPE TYROSINE KINASE-LIKE MOLECULES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Ave
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 KB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: ASCII/wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/335,865J
; FILING DATE: 19-January-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00210
; FILING DATE: 10-May-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL2358
; FILING DATE: 11-May-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6107472man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD-5277
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3100
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566
; TYPE: amino acid
; TOPOLOGY: linear
; ORIGINAL SOURCE: mouse
US-08-335-865J-9

Query Match 1.3%; Score 7; DB 3; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 PITSSSG 178
1111111
Db 250 PITSSSG 256

RESULT 14
US-08-335-865J-21
; Sequence 21, Application US/08335865J
; Patent No. 6107472
; GENERAL INFORMATION:
; APPLICANT: Stackner, Steven A.; Hovens, Christopher M.,
; APPLICANT: Wilks, Andrew F.
; TITLE OF INVENTION: RECEPTOR-TYPE TYROSINE KINASE-LIKE MOLECULES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Ave
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 KB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: ASCII/wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/335,865J
; FILING DATE: 19-January-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00210
; FILING DATE: 10-May-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL2358
; FILING DATE: 11-May-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6107472man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD-5277
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3100
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566
; TYPE: amino acid
; TOPOLOGY: linear
; ORIGINAL SOURCE: mouse
US-08-335-865J-21

APPLICANT: Wilks, Andrew F.
; TITLE OF INVENTION: RECEPTOR-TYPE TYROSINE KINASE-LIKE MOLECULES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Ave
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 KB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: ASCII/wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/335,865J
; FILING DATE: 19-January-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00210
; FILING DATE: 10-May-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL2358
; FILING DATE: 11-May-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6107472man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD-5277
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3100
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 593
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE: human
US-08-335-865J-21

Query Match 1.3%; Score 7; DB 3; Length 593;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 PITSSSG 178
1111111
Db 277 PITSSSG 283

RESULT 15
US-09-353-585-2
; Sequence 2, Application US/09353585
; Patent No. 6287865
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; Jones, David A
; Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses
; thereof
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 8th Floor, 1100 No. 6287865th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,585
FILING DATE: 15-Jul-1999
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q
1/68
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/930,277
FILING DATE: 27-OCT-1997
APPLICATION NUMBER: PCT/GB96/00785
FILING DATE: 01-APR-1996
APPLICATION NUMBER: GB 950658.5
FILING DATE: 31-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: MS. MARY J. WILSON
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-69
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1112 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Tomato
STRAIN: Cf2
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-353-585-2

Query Match 1.3%; Score 7; DB 4; Length 1112;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 NLSRLNL 204
11111111
Db 336 NLSRLNL 342

Search completed: July 17, 2002, 16:47:10
Job time: 145 sec

THIS PAGE BLANK (USPTO)


```

PF  biodiversity
XX
XX  Claim 20: SEQ ID NO 49154; 103pp; English.
PS
XX  The invention relates to isolated polynucleotide (I) and
XX  polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC  polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC  and gene mapping, and in recombinant production of (II). The
CC  polynucleotides are also used in diagnostics as expressed sequence tags
CC  for identifying expressed genes. (I) is useful in gene therapy techniques
CC  to restore normal activity of (II) or to treat disease states involving
CC  (II). (II) is useful for generating antibodies against it, detecting or
CC  quantitating a polypeptide in tissue, as molecular weight markers and as
CC  a food supplement. (II) and its binding partners are useful in medical
CC  imaging of sites expressing (II). (I) and (II) are useful for treating
CC  disorders involving aberrant protein expression or biological activity.
CC  The polypeptide and polynucleotide sequences have applications in
CC  diagnostics, forensics, gene mapping, identification of mutations
CC  responsible for genetic disorders or other traits to assess biodiversity
CC  and to produce other types of data and products dependent on DNA and
CC  amino acid sequences. ABG00010-ABG30377 represent novel human
CC  diagnostic amino acid sequences of the invention.
CC  Note: The sequence data for this patent did not appear in the printed
CC  specification, but was obtained in electronic format directly from WFO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 550 AA:

Query Match 69.2%; Score 1934.5; DB 22; Length 550;
Best Local Similarity 73.4%; Pred. No. 7.8e-193;
Matches 402; Conservative 9; Mismatches 30; Indels 107; Gaps 6;

QY 2 GATTGTGTRGVNDNRLSWSGSLAAPAAASSEMPELLAWSYFRRRKPOLCADLCTQ 61
   ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98 glpvtvgrqglftprplsawaapslaapaamsemepllawayfrirkqlcadlctg 157
QY 62 MLEKSPD-----QAAMILKARALTEMYTIDEIDVQEGIAEMLDENALQVPR 111
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 158 mlekspqdepdpelpyhqaawilkartlemvyidvdvqeglaemmldeala---- 213
QY 112 PGTSLKLPSTNQTGSPSOAVRPITQAGRPIPTQSGRPQSGRPQSGRPQSGRPQSGRP 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 214 -----pact--plqagrpitqglfipscqsgprpumegaitrplaytar 256
QY 172 PITSSGCFVRLGTASMLTSPDGPFINLSRLNLTKYSQKPLAKACLSISFIMKMLRL 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 257 plssagrfvrlg----- 269
QY 232 WIMWALSTENHSQYKDMWKKVIGKCYRRLGMYREAEKOFKSAKQOEKAVDTFLAKYV 291
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 270 -----rlgmyreaekqfksa|kxqgemvdfilylakyv 302
QY 292 SLDPVYALNLFKQGLDKFGEVTLGLGIARIYEMNNMNSAAEYKFEVLKODNTHYXAI 351
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 303 slpdpvcalnlfkqglkdfpgevlllglariyeemnmnsaaeykfevlkqdnthvcai 362
QY 352 ACIGSNHFSDQPEIALREFYRRLQNGIYNGQLFNNIGLCFVYAOQVDMTLTSEFALSL 411
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 363 acigsnhfysdpelalrfyrlllqmglyngqlfnniglcfcfyqgqdmcltstferalsl 422
QY 412 AENESEADVWYNLGHVAVGIGDTNLAHOCFRLLAVNNNNHAEYNNMLAVLEMRKG-HVE 470
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 423 aeneeseadvwnlghvavgidtnlahqclrlavvsnmnaeaynnlavlgdaegrhve 482
QY 471 QARALLQTASSIA-PHYEHPHFNFATISDKIGDQRSVVAOKSEAPPDHVDTOHLIKO 529
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 483 qarallqtassiscpyxktaflllqgslitfgldlqrsyvaakseaaifpdhvdcthlkq 542
QY 530 LROHFAML 537
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 543 lrqhfafl 550

```

```

RESULT 2
ID AAM41765
XX AAM41765 standard; Protein; 260 AA.
XX
XX AAM41765:
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 6696.
XX
XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0633450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSEQ-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX N-PSDB: AAI60921.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2: SEQ ID NO 6696; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with noctropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
SQ  Sequence 260 AA:

Query Match 40.8%; Score 1142.5; DB 22; Length 260;
Best Local Similarity 86.5%; Pred. No. 1.4e-110;
Matches 225; Conservative 8; Mismatches 24; Indels 3; Gaps 1;

```


XX (HXSE-) HXSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
PI N-PSDB; AAS82981.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 49153; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. Abg00010-Abg30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 207 AA:
SQ

Query Match 31.7%; Score 887; DB 22; length 207;
Best Local Similarity 85.0%; Pred. No. 5.1e-84; Matches 176; Mismatches 23; Indels 0; Gaps 0;
Conservative 8; Mismatches 23; Indels 0; Gaps 0;

QY 70 QAAWILKARALTEENVYIDEIDVDGEGIAEMMLDENAIQAVPRPGTSLKLPGTNGTGGPSQ 129
Db 1 gaaWilkaraltemvyideidvdgeglaemmldenaiqavprpgtslklpntngqpsq 60
QY 130 AVRPITTOAGRPITGFLRSTOSGPRGTMEQAIRPRATYARPTSSSGRFRVRLGTASML 189
Db 61 avrpitshakpiltgflrpsitgsgrpymegairprtarptlsssgfrvrlgtasml 120
QY 190 TSPDPPFINLSRLNLTYSOKPKLACLSISFIKKMMLRLIMLWLAISTESQYKKDMMW 249
Db 121 tspdpplfnxrlnltyksqpkmaakalpeylfihndvktalelaalstehsqykdww 180
QY 250 KVOIGKCYRILGMYREAEKQFSAALKQ 276
Db 181 kviglekryrilgmyreaeqqlsamkq 207

RESULT 5
ABG18793
ID ABG18793 standard; Protein: 142 AA.
XX
XX ABG18793;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #18784.
XX
XX Human, chromosome mapping; gene mapping; gene therapy; forensics;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PM
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PR
XX
XX (HXSE-) HXSEQ INC.
PA
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS82980.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 49152; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. Abg00010-Abg30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 142 AA:
SQ

Query Match 26.0%; Score 726; DB 22; length 142;
Best Local Similarity 100.0%; Pred. No. 1.9e-67; Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 YFRRKPOLCADLCTQMLEKSPYDQAAWILKARALTEENVYIDEIDVDGEGIAEMMLDENAI 105
Db 2 yfrrkfglccdlctcmlekepydqaawilkaraltemvyideidvdgeglaemmldena 61
QY 106 IAQVPRPGTSLKLPGTNGTGGPSQAVRPITTOAGRPITGFLRSTOSGPRGTMEQAIRPR 165
Db 62 iaqvprgtslklpntngqpsqavrpitgagrpiltgflrpsitgsgrpymegairpr 121
QY 166 TAYTARPTSSSGRFRVRLGTA 186
Db 122 taytarptlsssgfrvrlgta 142

RESULT 6
ABB68478
ID ABB68478 standard; Protein: 549 AA.
XX
XX ABB68478;

```

XX 26-MAR-2002 (first entry)
DT
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 32226.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL12581.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 32226; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL10840-ABL16175) and the encoded proteins
CC (ABB5737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX
SQ Sequence 549 AA:

```

```

Db 287 stgctlswwwqgmqrcllalypraeeflqgsfslfphdtylllsryyqrlkqpera 346
QY 300 LNFKQGLDKFRGEVYTLGLGARIVYEMNNMNSAAEYKKEVLKODNTHVXALACIGSNHF 359
Db 347 llvigeavdsrpfdivrlylegarlhqamegedalqlrlylaaklhpinvslaiaavgyf 406
QY 360 YSDQETALRFYRRLTOMGTYNGOLFNNLGGCCFYAAOYDWTLSFPRALSLANESEA 419
Db 407 ydmnpemalmyrrllslgaqspelycnialccllyggqldivlpcfqalataatqpgqs 466
QY 420 DWYNYLGHVAVGIDFTNLHQCFFRLALVNNNNHAEAVNNLAVLEMRKGHEQARALLQTA 479
Db 467 diwylsfvavtsgdfnlakrcqlcltsdaqngaalnlnlavlaagsgdllgaksylnaa 526
QY 480 SSLAPHMYE 488
Db 527 kdvmpdaae 535

RESULT 7
AAM3979
ID AAM39979 standard; Protein: 90 AA.
XX
XX AAM39979;
AC
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3124.
XX
KW Human; noctropic; immunosuppressant; cytosstatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW ankytrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000WO-US34263.
PF
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0532317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
XX
XX N-PSDB; AAI59135.
PS
XX
XX Example 4; SEQ ID NO 3124; 10078pp; English.
CC
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and

```


CC	XX	Query Match		6.9%; Score 194; DB 22; Length 1059;	
CC	XX	Best Local Similarity	21.3%; Pred. No. 2e-10;		
CC	XX	Matches 68; Conservative	67; Mismatches 140; Indels 44; Gaps		9
OY	253	IGKCYRRLGMYREAEKORFSALK-OOEWDFTPLYLAKVY-----SIDQPYTAL----	300		
Dd	124	Igvntkergqqlgealdnyrvrllppdfldgylmlaaalvaardmesavayltalqlynp	183		
OY	301	-----NLFK-----OGLDKPGEHVTLICGTARIYEEMNNKSSAAEYY	337		
Dd	184	dlyevrsdglgnllkalgrlleeakacylkaleetcpfagvawsnlgcvfnaggetlwaihnhf	243		
OY	338	KEVLKODMTHVAIACIGSNHFFYSOOPELALFRRLLQMGILYNGLFNNLGLCTCYAQ-	396		
Dd	244	ekavrlndfnldaylnlgnvlkearlfdravaaylrallnlsnnnavbnla-cvyyeqg	302		
OY	397	QYDMLTGFSEFRALSIAENEFEADWYNLGHAVGIGDPTNLAHOCPRALLVNNNNHAEAY	456		
Dd	303	lldaidcyrrateiqgn---fpdaycnlanalkckgykceadcyntalrlcshnadsl	359		
OY	457	NNLAVALENRKGGHEGARALLQTASSIAPHMYEPHFNFATISDKIGDLORS--YYAOKS	513		
Dd	360	nmlankireggyieatrilylkalvefpdfaahsnlasvldggqklkealmhykeatrl	419		
OY	514	EAAFPD-HVDTOHLIKOLR	531		
Dd	420	gpftadaysmngntlkeljg	438		
	RESULT_10				
ID	ABB67407	standard; Protein; 1059 AA.			
XX	XX	ABB67407;			
AC	XX				
DT	XX	26-MAR-2002 (first entry)			
XX	XX				
DE	XX	Drosophila melanogaster polypeptide SEQ ID NO 29013.			
XX	XX				
KW	XX	Drosophila; developmental biology; cell signalling; insecticide;			
KW	XX	pharmaceutical.			
OS	XX	Drosophila melanogaster.			
PN	XX	WO200171042-A2.			
PD	XX				
XX	XX	27-SEP-2001.			
PF	XX	23-MAR-2001; 2001WO-US09231.			
XX	XX				
PR	XX	23-MAR-2000; 2000US-191637P.			
XX	XX	11-JUL-2000; 2000US-0614150.			
PA	XX	(PEKE) PE CORP NY.			
PI	XX				
XX	XX	Venter JC, Adams M, Li PWD, Myers EW;			
DR	XX	WPI; 2001-656860/75.			
XX	XX	N-PSDB; ABL11510.			
PT	XX	New isolated nucleic acid detection reagent for detecting 1000 or more			
XX	XX	genes from Drosophila and for elucidating cell signalling and cell-cell			
PS	XX	interactions -			
XX	XX	Disclosure; SEQ ID NO 29013; 21pp + Sequence Listing; English.			
CC	XX	The invention relates to an isolated nucleic acid detection reagent			
CC	XX	capable of detecting 1000 or more genes from Drosophila. The invention			
CC	XX	is useful in developmental biology and in elucidating cell signalling and			
CC	XX	cell-cell interactions in higher eukaryotes for the development of			

Insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL6175) and the encoded proteins (ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1059 AA:

Query Match 6.9%; Score 194; DB 22; Length 1059;
Best Local Similarity 21.3%; Pred.No. 2e-10;
Matches 68; Conservative 67; Mismatches 140; Indels 44; Gaps 9;

Db

QY 253 IGRCTVRIKGMREAKRKORSAIK-QDEWMDTFLYLAKYY-----SILDQFVTAL---- 300
 | | ::||| ::||| | | | | |
Dd 124 Ignvkerqgqlqealdnyrravrlkpfdfldgylnlaaalvaardhesaqaylitalqynp 183
 | | |
QY 301 -----NLFR-----COIDKFPGSEVTLTGCIARIYEEMNNSSAAEY 337
 ||| ||| |
Dd 184 dlycvrsatlglllkylgrleekaeyklaleecpgfavaawslgcvfinaagelwiahnhf 243
 | | | | | | |
QY 338 KEVLKODNTHVXALACIGSNHFFYSDDPEIALNFYRRLDMGTYNQLFNNLGICCFYAQ- 396
 :: | | | | | | | | |
Dd 244 ekavefldpnfdlaylnlgnvlkearlfdraaaaylrallnspnavvhgnla-cvyeqg 302

QY 397 QYDMLTLTFERALLSLAENFEADAWYNILGHAVGIGDTNLAHOCFRLALVNNNHAELV 456
 | | | | | | |
Dd 303 lldlalddyyrtalfeiqpn---fpdaycnlanalkexgqkkeedcyntalrlicsnhadsl 359
 | | | | | | |
QY 457 NNLATLEMKRGHVDPARALLGTASSIAPHMYEPHFNFATISDKIGDLORS---YVAARKS 513
 ||||| :: | | | | | | | |
Dd 360 mlaiaikregqyleeatrilykalervfpdfaahnsiasvdqqgaklikealmhykeairl 419
 | | | | | | | | |
QY 514 EAAPPD-HVDTOHLIKQLR 531
 | | | | | | | | |
Dd 420 gptfadaysnmngntlkelq 438

RESULT 11

ABBB67408

XX ID ABB67408 standard; Protein; 1059 AA.
XX AC ABB67408;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 29016.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX DS Drosophila melanogaster.
PN WO200171042-A2.
PD 27-SEP-2001.
PP 23-MAR-2001; 2001WO-USO9231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
DR WPI; 2001-656860/75.
DR N-PSTDB; ABL11511.
PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell

Db 250 gdfvaltkyrvvacavpespplwnlmgcmfsgkkyaaaisclkranylapfdwk---1 306

QY 422 WYNIGHAVAGIDPNTLAHQCFRLALVNNNNHAEVNNLAVLEMKGHVEQARALLQTRASS 481

Db 307 lylglvhlvmqyasafhftisaainfqpkmgeylmlavaltnlledtenakrayaavh 366

QY 482 LAPHMEYEPHFNFATISDKIGLQRSYVAACKSE 514

Db 367 ldkcnpvlvnyavlllynggeknalagygeme 399

RESULT 13

AAW83960

ID AAW83960 standard; Protein; 519 AA.

XX AAW83960;

AC AAW83960;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 4166.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

XX EP1130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI, 2001-524255/58.

DR N-PSDB; AAK94927.

PT 830 Primers useful for synthesizing full length cDNA clones and their

PT use in genetic manipulation -

XX Claim 8; SEQ ID NO 4166; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA

CC clones. 830 cDNA molecules encoding a human protein have been

CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

CC molecules have been determined. Primers for synthesizing the full length

CC cDNA are useful for clarifying the function of the protein encoded by

CC the cDNA. The full length clones were obtained by construction of full

CC length enriched cDNA libraries that were synthesised by the oligo-capping

CC method. The primers enable the production of the full length cDNA easily

CC without any special methods. The present sequence is a polypeptide

CC encoded by a full length human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 519 AA;

SO

Query Match 6,58; Score 182.5; DB 22; Length 519;

Best Local Similarity 24158; Pred. No. 9,9e-10;

Matches 67; Conservative 49; Mismatches 148; Indels 9; Gaps 3;

QY 245 KDWMMKQVQIKCYRRLGMYREAEKQFKSALKQCEMVDPTFLYIAKVY---SLDQPTALN 301

Db 133 kdweismjvgylylqgfnkagqqlmalnlrhdltlylmjkinllegdlk---ale 189

QY 302 LFKOGIDKDFGEVTLGLGIARIYEENNMSSAAEYKEVLLKQDTHVXAIAICIGSNHFRYS 361

Db 190 vykkavetsfspentellctllglllylqlykafehlnalgydpcnykallaagsmmqth 249

QY 362 DQPEIALRFYRRLIQMKITVNGQLFNNILGLCCFVAQOYDMTLTSEKALSAENEEDADV 421

Db 250 gdfvaltkyrvvacavpespplwnlmgcmfsgkkyaaaisclkranylapfdwk---1 306

QY 422 WYNIGHAVAGIDPNTLAHQCFRLALVNNNNHAEVNNLAVLEMKGHVEQARALLQTRASS 481

Db 307 lylglvhlvmqyasafhftisaainfqpkmgeylmlavaltnlledtenakrayaavh 366

QY 482 LAPHMEYEPHFNFATISDKIGLQRSYVAACKSE 514

Db 367 ldkcnpvlvnyavlllynggeknalagygeme 399

RESULT 14

AAW82501

ID AAW82501 standard; Protein; 1151 AA.

XX AAW82501;

AC AAW82501;

DT 01-FEB-1999 (first entry)

DE C. elegans OGT protein.

XX C. elegans OGT protein.

DE OGT; O-linked GlcNAc transferase; uridine; transferase; tumour;

KW diphaspho-N-acetylglucosamine; polypeptide beta-N-acetylglucosaminyl;

KW predilection; type II diabetes; hyperglycaemia; Alzheimer's disease;

KW metastasis; diagnosis; inhibitor; treatment; diabetes mellitus.

XX Caenorhabditis elegans.

OS

XX Key Location/Qualifiers

PH Misc-difference 126

FT /label= unknown

XX W09844123-A2.

PD 08-OCT-1998.

XX 27-MAR-1998; 98WO-US06101.

PF 31-MAR-1997; 97US-0042270.

PR (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Hanover JA, Lubas W;

PI WPI; 1998-557118/47.

DR N-PSDB; AAV69302.

XX Protein exhibiting O-linked GlcNAc transferase activity, OGT -

PT useful, e.g. to assess predisposition to type II diabetes or

PT Alzheimer's or metastatic potential of tumours, and to identify

PT inhibitors

XX Claim 14; Page 39-42; 56pp; English.

XX This sequence represents a Caenorhabditis elegans OGT, O-linked GlcNAc

CC transferase protein (also known as uridine diphaspho-N-acetylglucosamine;

CC polypeptide beta -N-acetylglucosaminyl transferase). This protein is

CC useful to assess predisposition toward type II diabetes in patients

CC suspected of having hyperglycaemia that could evolve into this disease,

CC by assaying OGT activity in red blood cells. It can also be used to

CC assess predisposition toward Alzheimer's disease, to assess the

CC metastatic potential of tumours and to diagnose a tumour with metastatic

CC potential. OGT can also be used to identify OGT inhibitors, especially in

CC high-throughput assays, useful, e.g. in the treatment of diabetes

CC mellitus, tumour-derived diseases and Alzheimer's disease.

XX Sequence 1151 AA;

SO

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 16:42:24 ; Search time 15.46 Seconds

(without alignments)
848,418 Million cell updates/sec

Title: US-09-821-812-5

Perfect score: 2797
Sequence: 1 GGLTTTIGTIGRLGVDPRRLS.....PDHDTGHLIKQLRQHFAML 537Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents, AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
- 6: /cgn2_6/ptodata/2/1aa/6ackfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149.5	5.3	981	2	US-08-649-046-2
2	130.5	4.7	1079	3	US-09-058-489-22
3	130.5	4.7	1240	3	US-09-058-489-23
4	130.5	4.7	1347	3	US-09-058-489-24
5	103.5	3.7	452	1	US-08-336-618-23
6	103.5	3.7	560	1	US-08-336-618-22
7	101.5	3.6	580	3	US-08-906-865-1
8	101	3.6	438	2	US-08-897-340-34
9	101	3.6	438	4	US-09-252-329-34
10	99	3.5	308	2	US-08-897-340-33
11	99	3.5	308	4	US-09-252-329-33
12	99	3.5	484	2	US-08-879-260-4
13	99	3.5	484	3	US-09-231-529-4
14	99	3.5	484	4	US-08-977-816-4
15	95.5	3.4	459	1	US-08-336-618-12
16	95.5	3.4	459	1	US-08-336-618-26
17	95	3.4	308	2	US-08-807-050-3
18	95	3.4	1290	1	US-08-138-641-2
19	95	3.4	1290	1	US-08-138-133-2
20	94.5	3.4	292	2	US-08-879-260-2
21	93.5	3.3	319	5	PCT-US94-04208-2
22	92	3.3	308	2	US-08-807-050-1
23	92	3.3	513	1	US-08-200-232-4
24	92	3.3	513	5	PCT-US95-02219-4
25	92	3.3	513	5	PCT-US95-02219A-4
26	91.5	3.3	2471	3	US-09-112-450-4
27	89.5	3.2	1326	4	US-09-147-236-5

28	89.5	3.2	3031	1	US-07-689-008-2	Sequence 2, Appl
29	88.5	3.2	458	1	US-08-336-618-24	Sequence 24, Appl
30	87	3.1	1503	4	US-08-976-255-14	Sequence 14, Appl
31	86.5	3.1	1332	2	US-08-971-244-2	Sequence 2, Appl
32	86.5	3.1	1332	2	US-09-286-891-2	Sequence 20, Appl
33	85	3.0	521	3	US-08-996-338-20	Sequence 2, Appl
34	84.5	3.0	690	4	US-09-228-986-69	Sequence 69, Appl
35	83	3.0	843	4	US-09-235-451-25	Sequence 25, Appl
36	82	2.9	906	3	US-08-630-916A-48	Sequence 48, Appl
37	81.5	2.9	601	1	US-08-458-477A-2	Sequence 2, Appl
38	81.5	2.9	601	2	US-09-033-153-2	Sequence 2, Appl
39	81.5	2.9	601	4	US-09-325-430B-2	Sequence 2, Appl
40	81	2.9	309	3	US-08-996-338-22	Sequence 22, Appl
41	81	2.9	541	1	US-08-604-333-2	Sequence 2, Appl
42	81	2.9	541	3	US-09-110-618-2	Sequence 2, Appl
43	81	2.9	541	4	US-09-173-151A-28	Sequence 28, Appl
44	81	2.9	553	2	US-08-586-272-2	Sequence 2, Appl
45	81	2.9	553	3	US-09-082-969-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-649-046-2
; Sequence 2, Application US/08649046
; Patent No. 5912415
; GENERAL INFORMATION:
; APPLICANT: OLSZEWSKI, NEIL E.
; APPLICANT: JACOBSEN, STEVEN E.
; TITLE OF INVENTION: THE SPINDLE GENE, METHODS OF
; IDENTIFICATION AND USE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
; STREET: 119 NORTH FOURTH STREET, SUITE 203
; CITY: MINNEAPOLIS
; STATE: MINNESOTA
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,046
; FILING DATE: 16-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MCCORMACK, MTRA H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00340101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1225
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 981 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-649-046-2

Query Match 5.3%; Score 149.5; DB 2; Length 981;
Best Local Similarity 22.5%; Pred. No. 1.5e-07;
Matches 68; Conservative 45; Mismatches 90; Indels 99; Gaps 11;
OY 299 ALNLFKGLDLPFGVEVLT--CGIARIYEEEMNNMNSAAYKYKVKLQDNTHVAIAIC---353
Db 163 AFPCFSALRLDPRHNCALTHCGI--LHKREGRIVLEAASDYQKAMADASYKRAAEGLAI 220

[illegible]

```

RESULT      2
US-09-058-489-22
Sequence 22, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WH197-08PA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 22
LENGTH: 1079
TYPE: PRT
ORGANISM: Human
US-09-058-489-22

```

Query Match	4.78;	Score 130.5;	DB 3;	length 1079;
Best Local Similarity	18.58;	Pred. No. 2.7e-05;		
Matches	93;	Conservative	78;	Mismatches 205;
				Indels 127;
				Gaps 18

```

0Y 424 NUNHNVAVIGDTNLNHOEFLALVNNNNHAAVYNNLAVL-----EMKRGVNEOKRA- 474
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 SIGVLXOQONOPMDALGXUYICAYOLDJHNAAMAMMDLITVESCNOPODAKICYLNAARSK 379
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 475 -----LLQTAAS-----SLAPMYEPHNFATISDKIDDLORSVYAAQKSE 514
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 RCSNPTSLANIKFELQNGSDNMNNGQSLSHNPVOOVYSLCTPQKJDLHLEOLRANRDNLN 439
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 515 AAFRDNVDTOLIKOLRQNFAML 537
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 440 PA-----OKHOLELESQFVLM 456
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT      3
US-09-058-489-23
: Sequence 23. Application US/09058489
: Patent No. 6103886
: GENERAL INFORMATION:
: APPLICANT: Whitehead Institute for Biomedical Research
: APPLICANT: Lahn, Bruce
: APPLICANT: Page, David
: TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
: TITLE OF INVENTION: the Y Chromosome
: FILE REFERENCE: WH197-08PA
: CURRENT APPLICATION NUMBER: US/09/058,489
: CURRENT FILING DATE: 1998-04-10
: EARLIER APPLICATION NUMBER: 60/041,877
: EARLIER FILING DATE: 1997-04-11
: NUMBER OF SEQ. ID NOS.: 91
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 23
: LENGTH: 1240
: TYPE: PRT
: ORGANISM: Human
: US-09-058-489-23

```

Query Match	4.78;	Score 130.5;	DB 3;	Length 1240;
Best Local Similarity	18.58;	Pred. No. 3.5e-05;		
Matches	93;	Conservative 78;	Mismatches 205;	Indels 127;
				Gaps 18

QY 515 AAFPDHDTOLIKOLRHFAML 537
| : : : : :
Db 440 PA-----QKHOLEQLESOFVLM 456

RESULT 4
US-09-489-24
; Sequence 24, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; TITLE OF INVENTION: the Y Chromosome
; FILE REFERENCE: WH197-08PA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 1347
; TYPE: PRF
; ORGANISM: Human
US-09-058-489-24

Query Match
Best Local Similarity 4.7%; Score 130.5; DB 3; Length 1347;
Matches 93; Conservative 78; Mismatches 205; Indels 127; Gaps 18;

QY 97 AEMMDENALAOVPRPSTSLKLPCTNOGSGPSQAVRPITGAGRDTGFLRSTSGRPGT 156
| : : : : :
Db 19 AKKABEKGASRESESVSLVEERALEGMDSR-----LFGVRLHEDGARTKT 68

QY 157 -MCAITPTATATARTITSSSGFVRLGTASMLTSPDGPINLSRLNLTYSGPKPLAK 215
| : : : : :
Db 69 LIGGAVCYEGLIKAKGVEDFCQDGHNL-----LEDYSK----- 108

QY 216 ACTSISFIMKMLRLMTWLALSTHSQYK---DMW-----WKVOIGCYRLGMYREAEK 268
| : : : : :
Db 109 -----ALSAYQRYTSLQADYMKMAFLYGLGLVFTYNAFHHAIK 148

QY 269 QFSAKQOEVD-TFLYLAKVYVSLDOPVTALNLFKQGLDKPGEVTL-CG----- 319
| : : : : :
Db 149 AFQDVL---YDPSFCRAKEIHRLGIMFKVNTDYKSSLKHF--QLALIDCNPTLSMA 202

QY 320 -----IARIYEEMNNSSAAEYKEVLKQDN-----THXAIACIGSNHFTSD-----Q 363
| : : : : :
Db 203 EIOPHIAHLEETORKYSAEAYEQLLQTEMLPAQVATYVQDQGMHNNHNDLVGDRAIK 262

QY 364 PEIARFYRRLQMGITNGOLENNLGLCCFYAQOQYDMLTFSFEFALSIAEENEPADVY 423
| : : : : :
Db 263 ESAIYQYLOKSLSEADPRSGSWYFLGRCYSSIGKVDAFISYKOST---DSEESADTWC 319

QY 424 NLGHAVAGIGDTNLHOCFRLALVNNNNHAEVNNLAVL-----EMRKGHEQARA- 474
| : : : : :
Db 320 SIGVLYQQOQMDALQAYICAYOLDHGHAAAMMDLGTLYESCNOQDPAICYLNAARSK 379

QY 475 -----LLQTA-----SLAPMTEPHNEFATISDKIGDLQSYVAQKSE 514
| : : : : :
Db 380 RCSNTSTLAARIKFLQNGSDMMNGOSLSHHPOOVYSLCTPKTLOHLLEQLRANRDLN 439

QY 515 AAFPDHDTOLIKOLRHFAML 537
| : : : : :
Db 440 PA-----QKHOLEQLESOFVLM 456

RESULT 5
US-08-336-618-23

; Sequence 23, Application US/08336618
; Patent No. 5763590
; GENERAL INFORMATION:
; APPLICANT: Peattie, Debra A.
; APPLICANT: Harding, Matthew W.
; APPLICANT: Livingston, David J.
; TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
; TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
; TITLE OF INVENTION: CDNA
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,618
; FILING DATE: 09-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/963,325
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/777,752
; FILING DATE: 11-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/
; FILING DATE: 09-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 452 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-618-23

Query Match
Best Local Similarity 3.7%; Score 103.5; DB 1; Length 452;
Matches 79; Conservative 35; Mismatches 147; Indels 103; Gaps 14;

QY 179 RFVRLGTASMLTSPDGPINLSRLNLTYSGPKPL---AKKCLSTISFIMKMLRLMTWL 235
| : : : : :
Db 146 RRIKGVAKMGPPNDGANVEYA---LEGT-HKDRLETDQRELCFEYEGESIDLPAM--- 198

QY 236 ALSTHSQYKDMWKVQIGKCYRLGM-----YREAE-----KQFSALKQ 276
| : : : : :
Db 199 -----RRPFSAMKRRESLPSCSTSNLAMLAVWGRGRGSHHTAELRYEVRLKSFKAES 252

QY 277 QEWDTFLYLAKVYVSLDOPVTALNLFKQGLDKPGEVTLGLGIAARIYEEMNNSSAAEY 336
| : : : : :
Db 253 WEMSSARSMSRATYVK--ERGTA--YFREG-----K 279

QY 337 YKEVLKQDNTVHXAIACIGSNHFTSDQPEIARFYRRLQMGITNGOLENNLGLCCFYAQ 396
| : : : : :
Db 280 YKQALQYKKTIVSNLFTYSSSGEEMQKVHAKRLASHL-----NLAKCHLQLQ 327

QY 397 QYDMTLTFSFERALSIAEENEPADVYVNLGHAVAGIGDTNLHOCFRLALVNNNNHAEY 456
| : : : : :
Db 328 AFSAIIESCNKRALFELDSNNKGL---FRRGHAIHAVNDFDLARADFQKVLQLYPSNKA 384

QY 457 NNLAVLEMRKGVHEQARALLQTASSLADPHMYEPHNFATISDKIGDLQRSYVAQKSEAA 516
Db 385 TQLAVCOQR-----TRQLAREKKLYANMFE---RLAEEHKV-----KAEYA 424
QY 517 FPDH 520
Db 425 AGDH 428

RESULT 6

US-08-336-618-22
Sequence 22, Application US/08336618
Patent No. 5763590
GENERAL INFORMATION:
APPLICANT: Peattie, Debra A.
APPLICANT: Harding, Matthew W.
APPLICANT: Livingston, David J.
TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,618
FILING DATE: 09-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/963,325
FILING DATE: 16-OCT-1992
APPLICATION NUMBER: US 07/777,752
FILING DATE: 11-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/
FILING DATE: 09-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: YP191-06A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-618-22

Query Match 3.7%; Score 103.5; DB 1; Length 560;
Best Local Similarity 21.7%; Pred. No. 0.011;
Matches 79; Conservative 35; Mismatches 147; Indels 103; Gaps 14;

QY 179 RFRVLGTASMLTSPDGPINLSRLNLTYSOKPKL---AKACLSISFLMKMLRLLTWTL 235
Db 146 RRIRLVYKAMQGPNDGAVVEVA---LEGY-HKDRLELDQRELCFVGESESLDLPICAW--- 198
QY 236 ALSTEHSGYKRWMMKVQVGKCYRRLGM-----YREAE-----KOKRSALKQ 276
Db 199 -----RRPFSAMRKESIPSCSTNMLLAVWGRGSRSHRTAELRYEVRLKSEKAKES 252

QY 277 OEMVDTFLYAKVYVSLDQPYTALNLFKQGLDKFPEGEVTLTGCIARIYEENNNSSAAEY 336
Db 253 WEMSSARMSRATYVK--ERGTA--YFKEG-----K 279

QY 337 YKEVLKQDNTYHAXALACIGSNHFYSQPEIALRFRRLQWQITNGOLENNGLCCFYAQ 396
Db 280 YKQALLQYKKTIVSWLEYESSFSGEEMQVHALRLASHL-----NLAMCHLKLQ 327

QY 397 QYDMTUTSFERALSLAENDEEADYVYNGHVAVGIDPTNLAHQCFRLALVNNNHABAY 456
Db 328 AFSAIIESCKNALELDNSNNEKGL---FRGEAHLAVNDFDLARADFOVYLQLYPSNKAQ 384

QY 457 NNLAVLEMRKGVHEQARALLQTASSLADPHMYEPHNFATISDKIGDLQRSYVAQKSEAA 516
Db 385 TQLAVCOQR-----TRQLAREKKLYANMFE---RLAEEHKV-----KAEYA 424

QY 517 FPDH 520
Db 425 AGDH 428

RESULT 7

US-08-906-865-1
Sequence 1, Application US/08906865
Patent No. 6040168
GENERAL INFORMATION:
APPLICANT: Greengard, Paul
APPLICANT: Porton, Barbara
APPLICANT: Kao, Hung-Teh
TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hacksack Ave, Continental Plaza, 4th
STREET: Floor
CITY: Hacksack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,865
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-202
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-343-1684
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc = "Synapsin III"
HYPOTHETICAL: NO
ORGANISM SOURCE:
ORGANISM: Homo sapiens
US-08-906-865-1

Query Match 3.6%; Score 101.5; DB 3; Length 560;
Best Local Similarity 21.8%; Pred. No. 0.02;

OY 460 AYLEMRKGHEQA 472
DB 291 GTVNSKLRKLDDA 303

RESULT 14

US-08-977-816-4
; Sequence 4, Application US/08977816
; Patent No. 6194186
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEIN KINASE AND KINASE INHIBITORS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,816
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0429 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 484 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNORT11
; CLONE: 701698
; US-08-977-816-4

Query Match 3.58; Score 99; DB 4; Length 484;
Best Local Similarity 21.3%; Pred. No. 0.028;
Matches 54; Conservative 42; Mismatches 103; Indels 54; Gaps 9;

OY 260 LGMVREAEKQKFSALKQOEMVDF-----LYLAKYVSLDQPVTLNLFKQGLDFPGEVY 315
DB 65 LGRREA---LGDQOQSVRLDDSFVRGHLRGKCHLSLGNMAACRSFQALE----- 114
OY 316 LLLGCIARIYEEMNNMSAAEYK-----EVLKOD-----NTHVXAIA 352
DB 115 LDHNNAAQOQEFKNANVMYEKIAETDFEKDRPKYVFCMDRALERAPACHREKILAE 174
OY 353 CIGSNHYSDQPEIARFRLQMGITNGQLFNNLGLCCFYAAQYDMTITSEFASLSA 412
DB 175 CIAMIGRYPEAQSVAA-----SDILRMDSTNADALYVRGLCIYEDCIEKAVGFYQALRMA 230
OY 413 ENEEADAVWNL-----GHVAVGIGDTNLAHOCFRAL-VNINN---HAETVNNL 459

DB 231 PDHEKACIACRNAAKAKKEDGNKAFREGNYKLAVELYTEALGIDPNNIKTNALCYNR 290
OY 460 AYLEMRKGHEQA 472
DB 291 GTVNSKLRKLDDA 303

RESULT 15

US-08-336-618-12
; Sequence 12, Application US/08336618
; Patent No. 5763590
; GENERAL INFORMATION:
; APPLICANT: Peattie, Debra A.
; APPLICANT: Harding, Matthew W.
; APPLICANT: Livingston, David J.
; TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
; TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,618
; FILING DATE: 09-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/963,325
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/777,752
; FILING DATE: 11-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/
; FILING DATE: 09-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: VP191-06A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 459 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-336-618-12

Query Match 3.48; Score 95.5; DB 1; Length 459;
Best Local Similarity 23.18; Pred. No. 0.064;
Matches 43; Conservative 25; Mismatches 79; Indels 39; Gaps 6;

OY 355 GSNHYSDQPEIARFRLQMGITNGQLFNNLGLCCFYAAQYDMTITSEFASLSA 399
DB 277 GTVFEKGRKYQALLQYKKIYSLWLEYSSPSEDAQKAQALRLASHLNTLACHKLDQAFS 336
OY 400 MTLTSFERALSLAENEEAADWYNLGHVAVGIGDTNLAHOCFRALVNNNNHAEAVNML 459
DB 337 AAIESCKNALELDNNKGL---FRGPAHLAVNDFELARADFOKVLQVLPNNKAATOL 393
OY 460 AYLEMRKGHEQAALLQTAASSLAPHMYEPHFNFATISDKIGDLQKRYVAQKSEAAFPD 519

Db 394 AVCOOR-----IRROLAREKKLYANMF-----RIAEENK-----AKAEASSGD 433
QY 520 H-VDTQ 524
| | |
Db 434 HPTDTE 439

Search completed: July 17, 2002, 16:44:44
Job time: 140 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 16:42:49 ; Search time 20.96 Seconds
(without alignments)
2461.831 Million cell updates/sec

Title: US-09-821-812-5

Perfect score: 2797

Sequence: 1 GGLTFTVIGTRLGVDPRRLS.....PDHVDYQHILKQLRQHFMAL 537

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-71:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	950.5	34.0	458	2	T29520	hypothetical prote
2	197.5	7.1	806	2	AS3256	nuclear protein bl
3	194	6.9	379	2	F69210	conserved hypotet
4	193.5	6.9	334	2	F64399	hypothetical prote
5	186	6.6	403	2	B69196	conserved hypotet
6	185.5	6.6	1036	2	T33673	N-acetylglucosamin
7	185.5	6.6	1102	2	T39943	hypothetical prote
8	185	6.6	1194	2	E88499	protein K0467.3 [1
9	167	6.0	1173	2	T42719	TPR-containing/SH2
10	165.5	5.9	305	2	H97323	TPR-repeat-contain
11	160.5	5.7	774	2	T03919	hypothetical prote
12	158.5	5.7	1115	2	B84476	probable TPR repen
13	150.5	5.4	787	2	T16901	hypothetical prote
14	148	5.3	206	2	H69978	conserved hypotet
15	147	5.3	802	2	C84733	probable O-GlcNAc
16	146.5	5.2	1009	2	AF2033	hypothetical prote
17	145.5	5.2	1390	2	T14004	trfa protein - sli
18	145	5.2	584	2	AH2505	ser/chir protein ki
19	143.5	5.1	824	2	I52835	R-NOC - human
20	143	5.1	948	2	S75991	hypothetical prote
21	142	5.1	429	2	S77528	hypothetical prote
22	142	5.1	808	2	G86185	hypothetical prote
23	141.5	5.1	370	2	F84151	response regulator
24	140.5	5.0	339	2	B86577	hypothetical prote
25	140.5	5.0	339	2	C72048	tpi repeats-cl683
26	140	5.0	1432	2	S58819	antiviral protein
27	138.5	4.9	761	2	G70393	conserved hypotet
28	137.5	4.9	395	2	A12412	hypothetical prote
29	136.5	4.9	335	2	B71485	probable tpi repen

30	136	4.9	311	2	AE2353	hypothetical prote
31	135.5	4.8	707	2	AE2277	serine/threonine k
32	135.5	4.8	966	2	S25365	CYC8 protein - yea
33	134.5	4.8	393	1	C71371	conserved hypotet
34	134.5	4.8	478	2	H71261	conserved hypotet
35	133	4.8	256	2	G97031	TPR-repeat contain
36	133	4.8	402	2	A12030	hypothetical prote
37	133	4.8	469	2	A71322	hypothetical prote
38	132	4.7	665	2	A30185	nuclear protein nu
39	131.5	4.7	1119	2	B70126	surface-located me
40	130.5	4.7	1347	2	T02214	ubiquitous TPR mot
41	130	4.6	758	2	S45825	cell division cont
42	130	4.6	824	2	I49564	polycystic kidney
43	129.5	4.6	342	2	E70463	conserved hypotet
44	129.5	4.6	376	2	A96951	TPR-repeat contain
45	129.5	4.6	499	2	AE2278	hypothetical prote

ALIGNMENTS

RESULT	1	ALIGNMENTS
T29520	hypothetical protein T29520.5 - Caenorhabditis elegans	
C:Species: Caenorhabditis elegans		
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999		
C:Accession: T29520		
R:Pauley, A.; Gatlung, S.		
A:Submitted to the EMBL Data Library, July 1996		
A:Description: The sequence of C. elegans cosmid T29520.		
A:Reference number: 220634		
A:Accession: T29520		
A>Status: preliminary; translated from GE/EMBL/DBJ		
A:Molecule type: DNA		
A:Residues: 1-458 <PAU>		
A:Cross-references: EMBL:U64856; PIDN:AA04988.1; GSPDB:GN00023; CESP:T29520.5		
A:Experimental source: strain Bristol N2; clone T29520		
C:Genetics:		
A:Gene: CESP:T29520.5		
A:Map position: 5		
A:Introns: 55/2; 107/3; 181/2; 210/3; 271/3; 329/2; 388/1		
Query Match	34.0%	Score 950.5; DB 2; Length 458;
Best Local Similarity	40.1%	Pred. No. 1.2e-69;
Matches 182; Conservative 114; Mismatches 155; Indels 3; Gaps 3;		
Oy	78	RALTEMYIDEIDVDQEGIAEMMLDENAIAGVPRGTSCLKPGTNQCGPSQAVRPITQA 137
Db	2	QCLSDSTIVDELENDMGIAETFLDQNYIARNRGTGFARPKTS-AKGVNPILRPITNA 60
Oy	138	GRPTGFLRPSTGSGRPGTMOAIRTPRTAYTARITSSGRFVRLGASMLTSPDPPT 197
Db	61	GRPLSGVVRPQS-SFKSSGMDQAVTARTAKTARVSTASARNMRLGASMAAGDGEFV 119
Oy	198	NLSRLNLTYSQKRLAACLSTFTMKMRLIMLIMLALSTESQYDKMMKVOIGCY 257
Db	120	NLARLNTDKYADADPVNQLEFEYFYLYLNDIRVAHQIAGTSKAGFEDYWKQNLACY 179
Oy	258	YRLGAYREARQKOPFSAALKQOEAMDFELAYVYSLDOPVALNLFKGLKFPGEVTL 317
Db	180	LRLGMLDPAATQLOSLSEQKLIETFLALSAIRNVDPMAALKYSAGLEVPENVTML 239
Oy	318	CGIARIYEMNNMSAAYREYEVKQDNTHYXAIACISNHFYSQDPEIALRFRRLQM 377
Db	240	TGMARVQCALGEYDESVLTKRVLDAESNNIEALICAVTYYGGKPELARYYRIQM 299
Oy	378	GIYKQLFNNIGLCFYAQVDMTLTSFERALSLAENEEADVYNIQHAVCGIDNLT 437
Db	300	GVSSELEFLNTGLCCMAQDFALSTILRAQS-TMTDVAADVYNIQGLIVDGDVLS 358
Oy	438	AHOCRLALVNNNNHAEYNNIYAVLEMRKGVEQARALLOTASSIAPMYPPHFATIS 497

Db 359 AARSRIALSDPDHSESLVNLGITKHEGCKIDEARSLYSATSNPWFEGSNVLGLVS 418
Qy 498 DKIGDLORSYVAQKSEAFDPDHTOHLIKOLR 531
Db 419 FTGKYHCBRELIIEKALAAFPHEHCKILNLK 452

RESULT 2

A53256

nuclear protein bima - Emericella nidulans

C:Species: Emericella nidulans, Aspergillus nidulans

C>Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 21-Jul-2000

C:Accession: A53256; S21860

R:O'Donnell, K.L.; Osmanli, A.H.; Osmanli, S.A.; Morris, N.R.

J. Cell Sci. 99, 711-719, 1991

A:Title: bima encodes a member of the tetratricopeptide repeat family of proteins and is

A:Reference number: A53256; MUID:92121243

A:Accession: A53256

A:Molecule type: mRNA

A:Residues: 1-806 <OND>

A:Cross-references: EMBL:X59269; NID:g2334; PIDN:CAA41959.1; PID:g2335

C:Genetics:

A:Gene: bima

C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat

C:Keywords: mitosis; nucleus; tandem repeat

F:513-546/Domain: tetratricopeptide repeat homology <TT1>

F:547-580/Domain: tetratricopeptide repeat homology <TT2>

F:581-614/Domain: tetratricopeptide repeat homology <TT3>

F:615-648/Domain: tetratricopeptide repeat homology <TT4>

F:649-682/Domain: tetratricopeptide repeat homology <TT5>

F:683-716/Domain: tetratricopeptide repeat homology <TT6>

F:717-750/Domain: tetratricopeptide repeat homology <TT7>

F:751-784/Domain: tetratricopeptide repeat homology <TT8>

Query Match 7.1%; Score 197.5; DB 2; Length 806;

Best Local Similarity 21.6%; Pred. No. 9.2e-08;

Matches 101; Conservative 67; Mismatches 210; Indels 89; Gaps 12;

Qy 108 QVPRGTSLLKLPNTQOTGSPQAVRPITQACRPITGFLRPSTQSGRPTEQAIPTPTA 167
Db 357 QVAPHSVP---HSTDGQGGQRSVRLFNDI-KPSTNKISSRVALGKEBREVKKVMT--TG 410
Qy 168 YTAAPTSSS-GRVY---RLGTASMLTSPDGPFTNLRLNLTKYSQPKLA-----KAC 217
Db 411 NKARTTSSNNGRVVSGNNRRHAGEIHDGSKETSTSGSQNASKLAISERTKSV 470
Qy 218 LSIPIKMMRLIMLWAL-----SFEHSQYDMWKVKVQIGCYEELGYR 264
Db 471 EALWITLDFKIASGIFCLSRKCSDAIQIFSSLSGQOREPWLADIGRAYEQAMYT 530
Qy 265 EAERQF-----KSALKQDEAVDTFLYLAQVYSLDQVYALNLFKGLDKFGEVTL 317
Db 531 EAEKFFVAVKMAPSRLEDMETYSVLWHLKNDVL----- 566
Qy 318 CGIARIYEEMNNSSAAYKYEVKLKQDNTHYXALACISNHIFSDQPIALREFRLLIOM 377
Db 567 -----AYLAHELDVDRLSPCAWCAVNSFSHQDHDOALKCFERATOL 610
Qy 378 GIYNGQLFNNLGLCCFYAQOYDMTLTSPERALSLENEEADVYVNLGHAVGIDTNL 437
Db 611 DPHRAYGTTLOGHEVYANEEYDKALDAYRSGINDSRHYNA---WYGLGTYIDKKKGLDF 667
Qy 438 AHQCFRLALVNNNHAEVYNNLAVALERKGVGEQARALLQASSLAPMPYEPHFATIS 497
Db 668 AEQHRNAKINPNAVLICIGIVLEKMNPKSALLIYNNACILAPISVLAARPKAKAL 727
Qy 498 DKIGDLORSYVAQKSEAFDPDHTOHLI-----KQLR-----OHF 534
Db 728 MKLQDLKSALTTELKVLKDMADPDDEANVHYLLKLYKMLRDKGNALIKHF 774

RESULT 3

F69210
conserved hypothetical protein MTH83 - Methanobacterium thermoautotrophicum (strain D
C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jan-2000
C:Accession: F69210

R:Smith, D.R.; Doucette-Stamm, L.A.;

; Liu, D.; Spadafora, R.; Vitale, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,

K.I.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu

A:Reference number: A69000; MUID:98037514

A:Accession: F69210

A>Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-379 <MTH>

A:Cross-references: GB:A600079; GB:A600066; NID:g2621112; PIDN:AA84589.1; PID:g262

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH83

C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat

F:16-49/Domain: tetratricopeptide repeat homology <TT01>

F:50-83/Domain: tetratricopeptide repeat homology <TT02>

F:84-117/Domain: tetratricopeptide repeat homology <TT03>

F:118-151/Domain: tetratricopeptide repeat homology <TT04>

F:152-185/Domain: tetratricopeptide repeat homology <TT05>

F:186-219/Domain: tetratricopeptide repeat homology <TT06>

F:220-253/Domain: tetratricopeptide repeat homology <TT07>

F:254-287/Domain: tetratricopeptide repeat homology <TT08>

F:288-321/Domain: tetratricopeptide repeat homology <TT09>

F:322-355/Domain: tetratricopeptide repeat homology <TT10>

Query Match 6.9%; Score 194; DB 2; Length 379;

Best Local Similarity 21.9%; Pred. No. 5.9e-08;

Matches 70; Conservative 58; Mismatches 147; Indels 44; Gaps 5;

Qy 245 KDW-----WMKVQIGKYRYLGMTREAEKQFSALKQOEWDTFL-YLAQVYSLDQPV 297
Db 8 KDWIEGRAGWHLAGRSSLSKQGRKYLKAEFRKALKAPNPPEILHYNATLTLKLRPE 67
Qy 298 TALNLFKQGLDKFPEVYLLGCIARIYEEMNNSSAAYKE-----Y 340
Db 68 KALECTEYKILKNPKLAEAMNNKGLVLEKLGHYDEALCEYKALKINKRYAGAMNNKALV 127
Qy 341 LKQDNTHYXALACI-----GSNHFYSDQPIALRYRRLIOMGIYNGQ 383
Db 128 LKEIGRYDEALECYEKALQINPKLADAWYNNKGSVLIYLYKKYKALKCEKAIELNPKVNR 187
Qy 384 LFNNLGLCCFYAQOYDMTLTSPERALSLENEEADVYVNLGHAVGIDTNLAHQCFR 443
Db 188 AMGTGKITLHNLKIYEALKCYDVLQINPDOKA---WNKGLVFNELGRYDESLCEYE 244
Qy 444 LALVNNNHAEVYNNLAVALERKGVGEQARALLQASSLAPMPYEPHFATISIKIDL 503
Db 245 KALQINPKLAEAMNNKGVYLSLEGRYDEALCEYKALEIDPDDKTYMNNKGLVLEELKY 304
Qy 504 QRSYVAQKSEAFDPDHTOHLI-----KQLR-----OHF 534
Db 305 KDALECFQKALEINPEFAD 323

RESULT 4

F64399

hypothetical protein M0798 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: F64399

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

QY 511 OKSEAFPPD 519
Db 378 IRISPTFAD 386

RESULT 7

T39943
hypothetical protein SPBC236.09 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39943
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Pohl, T.
Submitted to the EMBL Data Library, May 1998
A:Reference number: 221892
A:Accession: T39943
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1102 <MOO>
A:Cross-references: EMBL:AL023287; PIDN:CA18877.1; GSPDB:GN00067; SPDB:SPBC236.09
A:Experimental source: strain 972h-; cosmid c2366
C:Genetics:
A:Gene: SPDB:SPBC236.09
A:Map position: 2

Query Match 6.6%; Score 185.5; DB 2; Length 1102;
Best Local Similarity 19.8%; Pred. No. 1.4e-06;
Matches 121; Conservative 84; Mismatches 209; Indels 197; Gaps 29;

QY 109 VPRCTSLKLGCTNO--TGGPSQAVRPTQAGR---PTTGFIRESTQSGRGTMEQAIRT 163
Db 217 VSQNGTYSIPAVNHPMAGOPIALAPVAPNQALPPIPQALPA--NGTENTLASPVTL 274
QY 164 PRTAVTAR-----PTSSGRFY-----RLGTASMTLS-----PDGP----- 195
Db 275 PAASAVQNAQVFWTSSPAAVVPQNTAATSTILAAOGAANLPNAPESYRHLISINE 334
QY 196 --FNLNLSL-----NLTKYS---QKPLAKACLSIFIMK---MMRLRLM----- 232
Db 335 ETWIDIGRLAELEFDQDKALSAVESALRQNPYSIPAMQIATILNRQFPALIEITYGT 394
QY 233 -----WLA-----LSTENSOYKD-----MMWKVOIGCYRL 260
Db 395 LDCPKQGEIWSALGHCYLMQDDLSRAVSARQALYHLKDKPKLWY---GIGLIDRY 451
QY 261 GMYREAEQKFSALKQ---OEMVDFTLYLAKVYSLDPYALNLFGQGLDKPFGVTL 316
Db 452 GSHEHAEEAFNOCCLMDPNFEKVNIEYRLGIITYKQHKFASQLELFRHILIDNPXPLTV 511
QY 317 L---CGIARIYEENNNSSAAEYKREVLKQDNTHVXACIG---SNHFFSDQPEIAL 368
Db 512 LDIYFOIGHVYEQREKYKLAKAEVERVLAETPNHAKVYLQGLWGLCHQSSSFTNO--DLAI 570
QY 369 REYRRL-----OMGIY-----NQGLNNNLGLCFY 394
Db 571 QYLTSLSEADDTDAOSWYLIGRCVYAAQKYNKAYEAYQAAVYRQGRNPTFECSTIGVLYQ 630
QY 395 AQQVDMTLTSPERALSIAEENEEADVYVNLGHV-----AVGIGDTNL 437
Db 631 INQVODALDAISRRL--NPISEVYVYDGLTYESCHNQISDALDYQRAAELEDPIN- 686
QY 438 AHQCFRLAV---NNNNH-----AEAYNNLAVLEM---RKGH 468
Db 687 PHIKARLQLLRGPNNEQKRIYNAPPSNVNPTAKYINQPGVPSNVPAQLSGNMOPRH 746
QY 469 VEQARALLQQTASS---LAPHYEP---HFNPATISDKIGDQRSVYAAQKSEAFPPHYD 522
Db 747 LPQAOPLPSATQSGVQVQPTQPSVTNNVAT--QPVIASTVVPQTAAFPSSQTAVPQTH 805
QY 523 TOHLIKQLRQH 533
Db 806 QSNAPFTPGKH 816

RESULT 8

E88499
protein K0AG7.3 [Imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88499
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99063613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: E88499
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1194 <STO>
A:Cross-references: GB:chr_III; PIDN:AA62535.1; PID:9687844; GSPDB:GN00021; CESP:K04
C:Genetics:
A:Gene: K0AG7.3
A:Map position: 3

Query Match 6.6%; Score 185; DB 2; Length 1194;
Best Local Similarity 22.3%; Pred. No. 1.7e-06;
Matches 113; Conservative 78; Mismatches 210; Indels 106; Gaps 21;

QY 93 QEGIAEMMLDENMA-----IAQVPRCTSLKLGCTNOGTGSPQAVRPTQAGRPTIGFLRP 147
Db 63 QQAIRTLFLTKINKYFSLINLTKGETS-----INSTNKVEA---TADNKSTIYSCQL 113
QY 148 STQ-----SGRPGTMEQAIRTPRTAYARTTSSGRFVRIG---TASMLT 190
Db 114 ENQAQQLAVANQQLPNLNATAVQQQLLTLPQOS--LAQILALAPQPTVVLNGVSEFLKKVT 172
QY 191 S-----PDGPFINLSR--LNLTKYSQKPKLAKCL--SISFLMKMLR--LMIWLALSTE 240
Db 173 ELAHRQFQSGVYVEAEKVCNLFVQSDPNMLPTLLLSAINFOTKMLEKSMOYSMLAIVN 232
QY 241 HSOYKDMWKKVOIGCYRRLGYRREAEKQFSAK-----OQEMVDFTLYLAKVYS 292
Db 233 N-----QCAEAYNLSNGIYKKEGQLODALENTYLAKEFEIDAIYNLAAIYS 282
QY 293 ---LDQPYTALNLFKQGLDKPGEVTLTGARIYEENNNSSAAEYKREVLKQDNTHVX 349
Db 283 GGDLEQAVTA---YENALQINPDLYCVASDGLNLLKAMGRLEAAKVYLKAIETQPOFAP 339
QY 350 AIAICGSNHFYSQPEI--ALRFYRLLQNGIYNGQLFNNIGLCCFYAAQVDMTLTSPER 407
Db 340 AMSNIGC---VENSQGEIMLAIHFEKAVTLDPNFLDAYINLGNVLEKARIFDRAVSAYLR 397
QY 408 ALSIAEENEE---EADAVYVNLGHVAVGIGDT-----NLAH----- 439
Db 398 ALNLSGNIAVYVGNLACYTYEGGLIDLAI--DTYKKAIDLQHPFPAVYCNLNAALKEKGSV 456
QY 440 ---QCFRLALVNNNNHNAEAYNNLAVLEMKGHEQARALLQTSASLAPHYEPHFNEAT 495
Db 457 VEADQMYKALELCPTHADSONNLANIKRQCKIEDATRYLKALEIYPERFAAHSNLAS 516
QY 496 I---SDKIGDQRSVYAAQKSEAFPPD 519
Db 517 ILQDQGLNDAILHYKEAIRIAPTFAD 543

RESULT 9

T42719
TPK-containing/SH2-binding phosphoprotein - mouse
N:Alternate names: 150TSP protein
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-May-2000
C:Accession: T42719
R:Malek, S.N.; Yang, C.H.; Earnshaw, W.E.; Kozak, C.A.; Desiderio, S.

J. Biol. Chem. 271, 6952-6962, 1996
A:Title: P150-TSP, a conserved nuclear phosphoprotein that contains multiple tetratricopeptide repeat motifs
A:Reference number: 206712; MID:56215125
A:Accession: T42719
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1173 <MAIL>
A:Cross-references: EMBL:LA9502; NID:g1236238; PID:g1236239; PIDN:AA642083.1
A:Experimental source: lymphoma
A:Genetics:
A:Gene: 150TSP
A:Map position: 7

Query Match 6.0%; Score 167; DB 2; Length 1173;
Best Local Similarity 18.3%; Pred. No. 4.9e-05;
Matches 114; Conservative 80; Mismatches 231; Indels 198; Gaps 18;
27 LAAPAAASSEMEPELLLWMSYFRRRKFDLCADICTOMLEKSPYDQ-----AW 73
Db 291 LAHAHFNTEVEAMQASCTQLARSHVQED-----YDAFQYIYQATQFASSTF 340
QY 74 ILKARALTEVYIDEIDVDQEGIAEMMLDENALAOVERPGTSLKPGTNDTGSPQAVRP 133
Db 341 VLPEFGIGQM-YI--YRGDKENASQCF--EKVLKAYPNNNYETMKILGSLVMAASDQEKRD 395
QY 134 ITQAGRTITGFLRSTQSGRGTEQALITPRTVITARPITSSGGRVRLGTASML---- 189
Db 396 IAK-----GHLKKVYQD-YPDDEAMVIEAQLIEO---TDIGALSANGTARRILOEK 444
QY 190 TSPDGFPIINLSRLNLTKYSQPKLAKACLSISFIMKMLRLIWLIALSTEHSGYKDMW 249
Db 445 VQADVP-----PEILN----- 455
QY 250 KVOIGKCYVILGMYREAEKQFSALKQOE-----AVDTFLYAKVYSLDQ 296
Db 456 --NVGALHFRNLGAEAKKFFLASLDRAKAEHDEHYNNAISVTSYNLARLYEAMCEF 513
QY 297 VTALNLEKQGLDKPRGVTLLCGIARIYEEMNNSSAAEYKKEVLKODNTVXALACGS 356
Db 514 HEAEKTLKNILREHPNVDCYLRIGAMARADGNFEASDWEKALQINQHPDAMSLIGN 573
QY 357 NHFSDOPETALRFYRRLDMGIVNGQLFNNILGICCFYAO-----QYDPTL 402
Db 574 LHLAKQWMPGQKKFERILNQPATQSPYSLMALGNWLOTLHQPTRDKREKRRQDRAL 633
QY 403 TSEFRAL-----SLAENEEDADVNLGHAANG 431
Db 634 AIYKQVLRNDAKNLUYAANGIGAVLAHKGFEARADVFAQVREARADISDVMNLAIHYE 693
QY 432 ----IGPTNLAHQCFRLALVNNNNHAEAYNNLAVLEMRKGVEDQARALLQTAASSLAPRM 487
Db 694 QKQYISAVOMENCLRRFYVHQNTVEVLY--LALALFKCGKLDQCKQTLARRHVAAPSDT 751
QY 488 EPHNFNFTI-----SDKIGDQFQSVVAOK 512
Db 752 VLMENVALVILQRLATSVLKDEKSNLKEVLANVKELELAHRYFSLKVGDMKRRDLALAA 811
QY 513 SEAAFPDHDVTQHLIKOLRQHPA 535
Db 812 SEAR-----QCSDDLISOAQYHVA 829

RESULT 10
H97323
TPR-repeat-containing protein [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H97323
R:Noiling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld

A:Reference number: A96900; MID:21359325; PMID:21359325
A:Accession: H97323
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <KDR>
A:Cross-references: GB:AE001437; PIDN:AAK81379.1; PID:g15026540; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
A:Genetics:
A:Gene: CAC3449

Query Match 5.9%; Score 165.5; DB 2; Length 305;
Best Local Similarity 25.0%; Pred. No. 9.2e-06;
Matches 56; Conservative 41; Mismatches 108; Indels 19; Gaps 5;
QY 274 LKQEDVDFTFLYLAQVYSLDQPTALNLFQGGDKPRGVTLLCGIARIYEEMNNSSA 333
Db 36 LTNGEDVDTIINALIYDLSCKVEKAEYKEALSIDDEERAVYGLATYIDYEDYEA 95
QY 334 AEYKKEVLKODNTVXALACGSMHFYSDQPEILARFYRRLLOMGIVNGQLFNNGLCCF 393
Db 96 IKLYNKALYIPNPHKAFELIANNVDSGKNLAIETYEKLISD--DMDFWSNLNLCCT 153
QY 394 YAOQ--YDMPLTSEFRALSLAENEEDADVNLGHAANGVIGDTNLAHQCFRLALVNNNN 451
Db 154 YEQNKNDLAVRLFSKALKIPNNHIAL---FNNGVICCKRMIEKALNFYEKSTENKKS 210
QY 452 HAEYNNILAVL---EMRKGVEDQARALLQTAASSLAPRMYPHP 491
Db 211 YEYSYLNLAIVYKHKDEKQ-----IQILSSGISCERTHF 246

RESULT 11
T03919
hypothetical protein F32D1.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: T03919
R:Becker, M.; Bradshaw, H.; Kramer, J.
Submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid F32D1.
A:Reference number: 215134
A:Accession: T03919
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-774 <BBC>
A:Cross-references: EMBL:AF016427; NID:g2291228; PIDN:AA65353.1; PID:g2291234
A:Genetics:
A:Map position: V
A:Introns: 47/2; 106/2; 135/1; 214/3; 391/2; 475/2; 666/3
A:Note: F32D1.3

Query Match 5.7%; Score 160.5; DB 2; Length 774;
Best Local Similarity 26.4%; Pred. No. 9.1e-05;
Matches 46; Conservative 36; Mismatches 79; Indels 13; Gaps 3;
QY 253 IGCYVRLGMYREAEKQFSALKQOEVDTFLYAKVYSLDQPTALNLEKQGLDKPRG 312
Db 593 MGTCTNELGDSQSAEKFGAIGENH-VNSYLTMALHILRONRSEVENLLRKAMTLAP 651
QY 313 EVTLGCIARIYEEMNNSSAAEYKKEVLKODNTVXALACGS-----NHFYSDQPEI 366
Db 652 SVTVLQNTALAEFHMQNTNRSLDFYRKALNDPHTLDSLOGTANLLOOTONHVESE--- 707
QY 367 ALRFYRRLLOMGIVNGQLFNNGLCCFYAODYDWTLLTSEFRALSLAENEEDADV 420
Db 708 --TFYRKVMEQAPNSYAHANVAGALVNLNQYDALAKVEYALALIDPDSVARE 759

RESULT 12
B84476
Probable TPR repeat nuclear phosphoprotein [imported] - Arabidopsis thaliana

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 16:43:39 ; Search time 13.46 Seconds

(without alignments)
1544.755 Million cell updates/sec

Title: US-09-821-812-5

Perfect score: 2797

Sequence: 1 GGLTFTVIGFRLGVDRRLS.....PDHVDVQHLIKQLRQHFAML 537

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt-40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	197.5	7.1	806	1 B1MA_EMENT	P17885 emericella
2	193.5	6.9	334	1 Y798_METJA	Q58208 methanococ
3	185.5	6.6	1036	1 OG11_RAT	P56558 rattus norv
4	161.5	5.8	920	1 OG11_HUMAN	O15294 homo sapien
5	150.5	5.4	787	1 YC091_CAEEL	P41842 caenorhabdi
6	143.5	5.1	824	1 CC27_HUMAN	P30260 homo sapien
7	138.5	5.0	761	1 YA88_AQUAE	O67178 aquifex aeo
8	138	4.9	1432	1 SK13_YEAST	P17883 saccharomyc
9	135.5	4.8	966	1 SSNE_YEAST	P14922 saccharomyc
10	134	4.8	665	1 NUCC2_SCHPO	P10505 schizosacch
11	130.5	4.7	1347	1 UTY1_HUMAN	O14607 homo sapien
12	130	4.6	758	1 CC27_YEAST	P38042 saccharomyc
13	130	4.6	824	1 TG37_MOUSE	O61371 mus musculu
14	128.5	4.6	824	1 TG37_MOUSE	O61371 mus musculu
15	127	4.5	612	1 PEX5_YEAST	P35056 saccharomyc
16	126	4.5	840	1 CC16_YEAST	P09798 saccharomyc
17	120.5	4.3	314	1 YD45_METJA	O58741 methanococ
18	120	4.3	403	1 YC0A_STNP7	P42460 synochococ
19	118.5	4.2	567	1 YE28_METJA	O58823 methanococ
20	117.5	4.2	296	1 PEX5_CANAL	O74711 candida alb
21	116	4.1	794	1 SEIL_HUMAN	O9ubv2 homo sapien
22	115.5	4.1	1333	1 UTX_MOUSE	O70546 mus musculu
23	115.5	4.1	1401	1 UTX_HUMAN	O15550 homo sapien
24	115	4.1	790	1 SEIL_MOUSE	O92249 mus musculu
25	114.5	4.1	1150	1 YKDI_CAEEL	Q03560 caenorhabdi
26	112.5	4.0	596	1 FLBA_CAUCR	P21296 caulobacter
27	109.5	3.9	468	1 YJ9H_YEAST	P47171 saccharomyc
28	108.5	3.9	472	1 IFM2_HUMAN	P09913 homo sapien
29	108	3.9	320	1 Y263_METJA	O06917 methanococ
30	107	3.8	1302	1 BCSC_ACEXY	P37718 acetobacter
31	106.5	3.8	289	1 SNAA_ARATH	O9spe6 arabidopsis
32	106.5	3.8	849	1 Y343_METJA	O57789 methanococ
33	105.5	3.8	508	1 KLC_DROME	P46824 drosophila

34	105.5	3.8	671	1 CUT9_SCHPO	P41889 schizosacch
35	105	3.8	626	1 CC23_YEAST	P16522 saccharomyc
36	105	3.8	1666	1 C1H_SCHPO	O10161 schizosacch
37	104	3.7	899	1 PRO6_YEAST	P19735 saccharomyc
38	103.5	3.7	256	1 YREC_STNP2	P19737 synochococ
39	103.5	3.7	318	1 SOLR_CLOAB	O9db5 mus musculu
40	103.5	3.7	619	1 KLC3_MOUSE	P33746 clostridium
41	103	3.7	327	1 ANX8_HUMAN	P13928 homo sapien
42	102	3.6	576	1 PEX5_PICPA	P33292 pichia past
43	101.5	3.6	560	1 SYN_HUMAN	O14934 homo sapien
44	101	3.6	375	1 RANP_BACSU	P45943 bacillus su
45	101	3.6	1447	1 B1RB_MOUSE	O9quk4 mus musculu

ALIGNMENTS

RESULT 1					
ID	B1MA_EMENT	STANDARD:	PRT:	806 AA.	
AC	P17885:				
DT	01-AUG-1992 (Rel. 23, Created)				
DT	01-AUG-1992 (Rel. 23, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Protein b1ma.				
GN	B1MA.				
OS	Emericella nidulans (Aspergillus nidulans).				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
OC	Eurotiiales; Trichocomaceae; Emericella.				
OX	NCBI_TaxID=5072;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SRRAIN-R153:				
MD	MDLINE=92121243; PubMed=1770001;				
RA	O'Donnell K.L., Osmani A.H., Osmani S.A., Morris N.R.;				
RT	"B1ma encodes a member of the tetrapeptide repeat family of				
RT	proteins and is required for the completion of mitosis in Aspergillus				
RT	nidulans.";				
RL	J. Cell Sci. 99:711-719(1991).				
CC	- FUNCTION: REQUIRED FOR THE COMPLETION OF MITOSIS IN ASPERGILLUS				
CC	NIDULANS.				
CC	- SUBCELLULAR LOCATION: Nuclear (Potential).				
CC	- SIMILARITY: BELONGS TO THE CDC27/NUC2 FAMILY.				
CC	- SIMILARITY: CONTAINS 8 TPR REPEATS.				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC					
DR	EMBL; X59269; CAA41959.1; -				
DR	PIR; S21860; S21860.				
DR	PIR; A53256; A53256.				
DR	InterPro; IPR001440; TPR.				
DR	PIfam; PF00515; TPR; 7.				
DR	SMART; SM00028; TPR; 7.				
KW	Cell division; Cell cycle; Mitosis; Repeat; TPR repeat;				
KW	Nuclear protein.				
FT	REPEAT 76 109 TPR 1.				
FT	REPEAT 127 160 TPR 2.				
FT	DOMAIN 260 399 B1MA DOMAIN.				
FT	REPEAT 513 546 TPR 3.				
FT	REPEAT 581 614 TPR 4.				
FT	REPEAT 616 648 TPR 5.				
FT	REPEAT 649 682 TPR 6.				
FT	REPEAT 684 716 TPR 7.				
FT	REPEAT 751 784 TPR 8.				
SO	SEQUENCE 806 AA; 89714 MW; F137BDE3A74C0A57 CRC64;				


```
Db 168 DQYNTALRCPTHADSLNNLANIKREOGNIEFAVRLYRALKLVEFEFAAHSHNLASVLCQ 247
Oy 395 AAOYDMTTFERRALSLAENEEDADWYNLGHVAVGIGDTNLAHQCFPLALVNNNNHAE 454
Db 248 OQKLOEALMHYKEAIRISPT---FADASNMKNTLKEKMDVGCALQCTTRAIQINPARAD 304
Oy 455 AYNNILAVLEMRKGVHQAARALIQTASSLAPHYEPHFENFATISDKIGDQSRSYAAKSE 514
Db 305 AHSNLIASIHKSGNIPERAIASRYTRALKLPDPPDAYCNLAHCLQIVCDYDERMKLV 364
Oy 515 AAFPDHVDQHILKQLRHFAWL 537
Db 365 STVADQLEKNRL-PSVHPHSHML 386

RESULT 5
YO9L_CAEEL STANDARD: PRT; 787 AA.
ID YO9L_CAEEL
AC P41842.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 90.0 kDa protein T20B12.1 in chromosome III.
GN T20B12.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitida; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN NCH1
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RL Waterston R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO YEAST YNL313C.
CC -1 SIMILARITY: CONTAINS 6 TPR REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
Db 429 RRLIQKPNDSMLHYLGDITRNLLEYFTKAIELSDDRNARARSIGHLIMDKKEFEAYK 488
Oy 404 SFERALSLAENEEDADWYNLGHVAVGIGDTNLAHQCFPLALVNNNNHAEYNNLAVLE 463
Db 489 HIRRSLEL---OPIQGTWENAGYCAWKLENKESTQCYHRCVSIQPDHFEMNLSAY 545
Oy 464 MRKGVHQAARALIQTASSL--APHYEPHFENFATISDKIGDQSRSYA 509
Db 546 IIRHGGKPRAMKRLQELALYNNENHFWME---NYMLLSVDYGFSAIQA 591

RESULT 6
CC27_HUMAN STANDARD: PRT; 824 AA.
ID CC27_HUMAN
AC P30260; Q16349;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein CDC27hs (Cell division cycle protein 27 homolog) (H-NUC).
GN CDC27.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN NCH1
RP SEQUENCE FROM N.A.
RX MEDLINE=94052097; PubMed=8234252;
RA Tugendreich S., Boguski M.S., Seidlin M., Hieter P.A.;
RT "Linking yeast genetics to mammalian genomes: Identification and
RT mapping of the human homolog of CDC27 via the expressed sequence tag
RT (EST) data base."
RL Proc. Natl. Acad. Sci. U.S.A. 90:10031-10035(1993).
RN NCH1
RP SEQUENCE FROM N.A.
RX MEDLINE=95275739; PubMed=7756179;
RA Chen F.L., Deng Y.C., Durfee T., Chen K.C., Yang-Feng T., Lee W.H.;
RT "Identification of a human homologue of yeast nucz which interacts
RT with the retinoblastoma protein in a specific manner."
RL Cell Growth Differ. 6:199-210(1995).
CC -1 SUBUNIT: INTERACTS WITH RB.
CC -1 SUBCELLULAR LOCATION: Nuclear.
CC -1 SIMILARITY: BELONGS TO THE CDC27/NUC2 FAMILY.
CC -1 SIMILARITY: CONTAINS 9 TPR REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L36940; AAA50573.1; .
 DR PIR: S29142; S29142.
 DR SGD: S0006393; SKT3.
 DR InterPro: IPR001440; TPR.
 DR Pfam: PF00515; TPR; 7.
 DR Nuclear protein; TPR repeat; Repeat.
 FT REPEAT 4 37 TPR 1.
 FT REPEAT 47 80 TPR 2.
 FT REPEAT 425 458 TPR 3.
 FT REPEAT 471 507 TPR 4.
 FT REPEAT 508 541 TPR 5.
 FT REPEAT 627 661 TPR 6.
 FT REPEAT 702 735 TPR 7.
 FT REPEAT 736 769 TPR 8.
 FT REPEAT 945 985 TPR 9.
 FT REPEAT 987 1018 TPR 10.
 FT REPEAT 1226 1259 TPR 11.
 SO SEQUENCE 1432 AA; 163737 MW; 29142C7D769FB70A CRC64;

Query Match 4.9%; Score 138; DB 1; Length 1432;
 Best Local Similarity 17.7%; Pred. No. 0.0077;
 Matches 112; Conservative 83; Mismatches 184; Indels 252; Gaps 25;

QY 46 YFRKKPOLCADLCTQMLEKSP-----YDQAMILIKALTEMYIDE-----IDVDQGI 96
 DB 518 FIERNMWMDATLTLQVHEQSPNNLEVLSELWM-----SKAHMGYMDALAGLDIVIGKI 572
 QY 97 AEMML-----DENALAQVPRPTSLKL-----PGTNO 123
 DB 573 KGMDLRSIDFRALNMRQAKYIMKHAINDAKQENVCAPKRLIQSKILDTFRPGEST 632
 QY 124 TG-----GPSQAVRPITQAGRPITGFLRPSTQSGRGTME 158
 DB 633 LGDIYCHYKDLRAFCYKFAFDLADGDTAAKYITE----- 670
 QY 159 QAIRPRATYARP-----ITSSSGFVRLGTA-SMLTSPDGFILSLNLTIKYSQAKRL 213
 DB 671 -----TYASKRPWQAASSIASRLNGERAKAKELRSNNMFRVVGIAHLEKQESDSI 722
 QY 214 AKACISIFIMKMLRLIMLALSTESQYKMMWKVOIGKCYRLGMYREAEKQFSA 273
 DB 723 E-----WFOALRVDPNDYESW---VGLQAYHACRIEASIKVDFKA 762
 QY 274 LKQ-----EMWDFLYLAKV-----YVSLDQV--TALMLF 303
 DB 763 IQLRPSHTFAQYFAISLGDVGEYLSLIDLEKVCQEAATEESPQIGLVEVLMRCSLDLY 822
 QY 304 KQG-----LDKPFGEVTL 316
 DB 823 SQGLFLKSVSIKADTERIKIITISELKCENQVWIIYSLVRLFTWISKVDTLPVE-- 879
 QY 317 LGLAIRITEENNNSSAAEYKEVLKQDNTHV-----XAIAC-----IGSNHFPYS 361
 DB 880 ---SLVSTIE--NSQESSE---EIDSVNIIKIDTLDDSTTDNNISACKPILILASKYSVS 932
 QY 362 DQ--PEIALRFYRLLLONGIYNGOLFNNLGLCCFYA-----QQYDMLTSPFERALS 412
 DB 933 DQKFDIA-----GTVRASVWYNIIGISELTAFTLKEPQYRDAIFAFAKKSIOLO 982
 QY 413 ENEEAAADVWNTGLHVAAGICIDTNLAHOCFRALVNNNNHAEVYNNLAVLEMRKHVDOA 472
 DB 983 SN---ISETWGLGATAMD--NFRVSQHCPIKATALEPKATNTWENLAMLGLKKDTEFA 1038
 QY 473 RALLQTAASLAPMHYEPHFNFATISDKIGDL 503
 DB 1039 QQVINKLQSLAPQDSSPWLGMALLIEGGDI 1069

RESULT 9
 ID SSN6_YEAST STANDARD; PRT; 966 AA.
 AC P14922;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1990 (Rel. 14, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Glucose repression mediator protein.
 GN SSN6 OR CYC8 OR YBR112C OR YBR0908.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89211664; PubMed=2854095;
 RA Trumbly R.J.;
 RT "Cloning and characterization of the CYC8 gene mediating glucose
 RT repression in yeast."
 RL Gene 73:97-111(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8805502; PubMed=3316983;
 RA Schultz J., Carlson M.;
 RT "Molecular analysis of SSN6, a gene functionally related to the SNF1
 RT protein kinase of Saccharomyces cerevisiae."
 RL Mol. Cell. Biol. 7:3637-3645(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=92327848; PubMed=1626431;
 RA Mannhaupt G., Stucka R., Ehnl S., Vetter I., Feldmann H.;
 RT "Molecular analysis of yeast chromosome II between CND1 and LYS2: the
 RT exsition repair gene RAD16 located in this region belongs to a novel
 RT group of double-finger proteins."
 RL Yeast 8:397-408(1992).
 RN [4]
 RP TPR REPEATS.
 RX MEDLINE=90124639; PubMed=2404612;
 RA Sikorski R.S., Boguski M.S., Goebel M., Hieter P.A.;
 RT "A repeating amino acid motif in CDC23 defines a family of proteins
 RT and a new relationship among genes required for mitosis and RNA
 RT synthesis."
 RL Cell 60:307-317(1990).
 CC -1- FUNCTION: IT IS INVOLVED IN REPRESSION BY A1-ALPHA2 AND ALPHA2 AND
 CC IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS
 CC PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT
 CC DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 10 TPR REPEATS.
 CC -1- SIMILARITY: TO YEAST GAL1 AND CCR4.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M23440; AAA34545.1; .
 DR EMBL: M17826; AAA35103.1; .
 DR EMBL: X66247; CAA46973.1; .
 DR EMBL: X78993; CAA55615.1; .
 DR EMBL: Z35981; CAA85069.1; .
 DR PIR: S25365; S25365.
 DR SGD: S0000316; CYC8.
 DR InterPro: IPR001440; TPR.
 DR Pfam: PF00515; TPR; 10.
 DR SMART: SMO0028; TPR; 9.
 KW Transcription regulation; Repressor; Repeat; TPR repeat;
 KW Nuclear protein.

```

FT DOMAIN 15 30 POLY-GLN.
FT REPEAT 46 79 TPR 1.
FT REPEAT 80 113 TPR 2.
FT REPEAT 114 147 TPR 3.
FT REPEAT 150 183 TPR 4.
FT REPEAT 187 220 TPR 5.
FT REPEAT 224 257 TPR 6.
FT REPEAT 258 291 TPR 7.
FT REPEAT 296 329 TPR 8.
FT REPEAT 330 363 TPR 9.
FT REPEAT 364 398 TPR 10.
FT DOMAIN 493 556 30 X 2 AA TANDEM REPEATS OF Q-A.
FT DOMAIN 557 587 POLY-GLN.
FT CONFLICT 547 547 K -> Q (IN REF. 3).
SQ SEQUENCE 966 AA; 107202 MW; 84B509CF3208C5C0 CRC64;

Query Match 4.8%; Score 135.5; DB 1; Length 966;
Best Local Similarity 22.5%; Pred. No. 0.007;
Matches 60; Conservative 36; Mismatches 114; Indels 57; Gaps 8;

QY 264 REAEKFKSLKQGEVDFLYLAKYVSLDQPYTALNLFKGLDKEF--PGEVTLICGIA 321
DB 98 RAELRYERLALVNPBLSDVWATLGHCIYMDLDQRAYNAYQALYHLNPNVPRILMHGIG 157
QY 322 RIYEENNNSSAAEYKVELKODNTVYXAIACIGSNHPSYDQPEILRLFRRLONGITN 381
DB 158 ILDRGSLDYAEAFRAKYLELD-----PHE-----EKANEITFRL----- 193
QY 382 GOLFNNLGLCCFYAQYDMLTFSFERALSLAENEEDADWYNIQHVAVGIDGTNLAHQC 441
DB 194 GLTYKHOG-----KMSQALECFRILPQPPAPLDQMDIMFQGLSGVLESMGEQAKKA 246
QY 442 FRLALVNNNNHAAVYNNLAVLEMRKGVHQAARLLQTAASLAHMYEPHNPATISXIG 501
DB 247 YEHTVLAQNOHNAKAVLQOLGCL-YGMSNVQ-----FDDP-----OKAL 282
QY 502 DLQRSYVAQKSEAFPPDHVDTOHLK 528
DB 283 DYLLKSLADPDDATTWYHLGRVHMIR 309

RESULT 10
NUC2_SCHPO STANDARD; PRT; 665 AA.
ID NUC2_SCHPO
AC P10505;
DT 01-JUL-1989 (rel. 11, Created)
DT 01-OCT-1996 (rel. 34, last sequence update)
DT 16-OCT-2001 (rel. 40, last annotation update)
DE Nuclear scaffold-like protein p76.
GN NUC2 OR SPAC17C9.01C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TS MUTANT NUC2-663;
RA MEDLINE=88198361; PubMed=3283148;
RA Hirano M., Hirooka Y., Yanagida M.;
RT "A temperature-sensitive mutation of the Schizosaccharomyces pombe
RT gene nuc2+ that encodes a nuclear scaffold-like protein blocks
RT spindle elongation in mitotic anaphase.";
RL J. Cell Biol. 106:1171-1183(1988).
RN [2]
RN RP REVISION TO 649.
RA Yanagida M.;
RL Submitted (MAR-1989) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-557 FROM N.A.
RC STRAIN=972;
RA Murphy L., McDougall R., Jones L., Simpson I., McNeill A., Harris D.,

```

```

RA Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP DOMAINS.
RX MEDLINE=90124640; PubMed=2297790;
RA Hirano M., Kinoshita N., Morikawa K., Yanagida M.;
RT "Snap helix with knob and hole: essential repeats in S. pombe nuclear
RT protein nuc2+ ";
RL Cell 60:319-328(1990).
CC -1- FUNCTION: NUC2 INTERACTS WITH SPINDLE APPARATUS, CHROMOSOMES,
CC OR NUCLEAR ENVELOPE, AND INTERCONNECT NUCLEAR AND CYTOSKELETAL
CC FUNCTIONS IN MITOSIS, SO THE ELONGATION OF THE SPINDLE IN ANAPHASE
CC IS BLOCKED.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE CDC2/NUC2 FAMILY.
CC -1- SIMILARITY: CONTAINS 9 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: X07693; CAA30532.1; -.
CC DR EMBL: Z73099; CAA97347.1; -.
CC DR PIR: A30185; A30185.
CC DR InterPro: IPR001440; TPR.
CC DR Pfam: PF00515; TPR; 8.
CC DR SMART: SM00028; TPR; 8.
CC KW SMART division: Cell cycle; Mitosis; Repeat; TPR repeat;
CC KW Nuclear protein.
CC FT REPEAT 115 148 TPR 1.
CC FT DNA_BIND 191 257
CC FT REPEAT 329 362 TPR 2.
CC FT REPEAT 363 396 TPR 3.
CC FT REPEAT 431 464 TPR 4.
CC FT REPEAT 466 498 TPR 5.
CC FT REPEAT 499 532 TPR 6.
CC FT REPEAT 534 566 TPR 7.
CC FT REPEAT 568 600 TPR 8.
CC FT REPEAT 601 634 TPR 9.
CC FT MUTAGEN 504 504 G->D: IN TEMPERATURE SENSITIVE MUTANT.
CC FT CONFLICT 440 440 C -> W (IN REF. 1).
CC SQ SEQUENCE 665 AA; 76171 MW; 97775DC061ECABFA CRC64;

Query Match 4.8%; Score 134; DB 1; Length 665;
Best Local Similarity 19.7%; Pred. No. 0.0054;
Matches 113; Conservative 69; Mismatches 245; Indels 148; Gaps 19;

QY 6 TWIGTRLGVDK-----PRLKMSAGPSLAAPAMSSEMEPLIL---AMSYFRRRKFOQLCAD 57
DB 156 TALGVPLDANNVFEVILPYPLTAMKGFKEKSGQTWATASVPEPFLKRSKSSSSSSNFVSSES 215
QY 58 LCTQMLEKSPYDOAAWILKARALTTEWYIDEIDVDQEGIAEMMDENAIQVPPGCT--- 114
DB 216 IANSYSNS-----ISAFPK--WFDKVDASE-----LPGSEKE 246
QY 115 ---SLKLPNTONGSPQSAVPIPTQACRP-----ITGFLRPSYSGRPGTMEQAIPTPR 165
DB 247 RHQSLKIQSOSQTSKNLLAFNDQAKAOSNNROTSLKSHFVPRQIALRPGA-----R 298
QY 166 TAYTARDPTSSS--GREYRLGTASMLTSPDGFINLSRL-----NLTKYSQKPIARACT 218
DB 299 LTYKLRBARSSKRGK---STPQSFREDNNLMELKLFKGGYVLLAQY--KLREALNCF 352
QY 219 SISIMKMLRLMILMIALSTEHSHQYKDWMMKQVIGKCYIYLGMYRAEKQFK----- 271
DB 353 Q-----SLPIEQONTPEVLALRIGITPELVYERKSEEVFOKRLDLS 394
QY 272 SALKQGEVDFLYLAKYVSLDQPYTALNLFKGLDKEFGEVTLICGIARIYEMNNMS 331

```

```

Db 395 SRVADMEVFSTALM-----HLQKSVPLSYLAHETLETPNPSPESSWCILACFSLQREHS 448
QY 332 SAAEYKEVELKQDNTVHVAACIGSNHFPYSDQPEIALRFYRLLQMGTYNCQLENNLGLC 391
Db 449 QALCINAIQIDLPTEFAVAYLQGEHSAN----- 478
QY 392 CFYAOQDMLTSEFRLSLAENEBEADAVYNIQHAVGIGDTNLAHQCFRLALVNNN 451
Db 479 -----EEVEKSKTSFRKAIKRVNVRHNA---WYGLGMVYLKTRNDQADFHQRAAEINPN 531
QY 452 HAAEYNNLAIVLEMKGHEQARALLQTTASSLAPHNVEPHFATISDKIGLOLSSVAAQ 511
Db 532 NSVILITGMIYERCKDYKALDFYDRCKLDEKSSLARFKKAVYLLHHDKALVELE 591
QY 512 KSEAAPDHVDTOHLI-----KQLR-----OHFAM 536
Db 592 QLKAIAPDEANVHFLKGIKFKQMKRKNLAKHFTI 626

RESULT 11
UTY_HUMAN STANDARD; PRT; 1347 AA.
AC 014607; 014608;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ubiquitously transcribed YPR protein on the Y chromosome).
GN Ubiquitously transcribed YPR protein on the Y chromosome).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98022381; PubMed=9381176;
RA Lehn B.T., Page D.C.;
RT "Functional coherence of the human Y chromosome.";
RL Science 278:675-680(1997).
CC - SUBCELLULAR LOCATION: Nuclear (Potential).
CC - ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC - SIMILARITY: CONTAINS 5 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000994; AAC51841.1; -
DR EMBL; AF000995; AAC51842.1; -
DR EMBL; AF000996; AAC51843.1; -
DR MIM; 400009; -
DR InterPro; IPR003347; JmJC.
DR InterPro; IPR001440; TPR.
DR Pfam; PF02373; JmJC; 1.
DR Pfam; PF00515; TPR; 7.
DR SMART; SM00028; TPR; 6.
KM Repeat; TPR repeat; Nuclear protein; Alternative splicing.
FT REPEAT 93 121
FT REPEAT 130 158
FT REPEAT 167 196
FT REPEAT 318 346
FT REPEAT 352 380
FT REPEAT 996 1079
FT VARSPLIC
FT
FT
FT
FT
FT
FT
FT
FT
FT
FT

```

```

SQ SEQUENCE 1347 AA; 149577 MW; C28B870127107E71 CRC64;
Query Match 4.7%; Score 130.5; DB 1; Length 1347;
Best Local Similarity 18.5%; Pred. No. 0.028;
Matches 93; Conservative 78; Mismatches 205; Indels 127; Gaps 18;

QY 97 AEMMLDENALIAQVPRPGTSLKLPGTNQTGSPSAVPRITTOAGRPITGFLRSTQSGRPGT 156
Db 19 AKMAEGKASRESESESVSLVEEREAALGMDSR-----LFGFVRLHEDGARFTKT 68
QY 157 -MEQALRPRIYATYARPTTSSGGRVRLGTASMLTSPGPTINLSRLNLTYISQPKLAK 215
Db 69 LLGRAVRCYSELILKAEKVESDFEQGLGHNNL-----LEDYSK----- 108
QY 216 ACLSISIFMKMMLRLMIWLSTEHGQYK---DMW-----WKVOIGKCYRLGYAREAK 268
Db 109 -----ALSAVQRTYSIQADYKKAAPLYGLGLVYFYNAFHMAIK 148
QY 269 QFKSALKQEMVD-TFLYLAKVYVSLDQPYTALNLFKQGLDKPGEVTL-CG----- 319
Db 149 AFQDVL-----YVDPSPCRAKELHLRLGLMFKVNTDYKSSLKHf--QLALIDCNPTLSNA 202
QY 320 -----IARIYEEAMNMSAAYKEVLAQDN-----THYXAIACIGSNHFTSD-----Q 363
Db 203 EIQEHIALHLYETQRYHSAKAYEDOLLQTEMLPAQVATVYLOQIGMHNNHNDLVGDKATK 262
QY 364 PEILAFYRRLRLQNGIYVGFNNIGLCFFAAQOYDWTITFEFRLSLAENEBEADAVY 423
Db 263 ESYAIOYLQKLEADPNQSGQWYFLGRCYSSIGKVOADFIYRSTI---DSEVASADTWC 319
QY 424 NLGHVAVGIGDTNLAHOCFRLALVNNNHAEVYNNLAVL-----EMRGHVEQARA- 474
Db 320 SIGVLYQQNQPMDALOAYICAVOLDHGHAAMMDGLTYESCNOPODAICYLNAARSK 379
QY 475 -----LLQTAS-----SLAPHMYEPHFATISDKIGLOLSSVAAQKSE 514
Db 380 RGSNTSTLAIRIKFLQNGSDMNGQSLSHHPVOQVYSLCTLPQKLOHLEOLRANRONLN 439
QY 515 AAFPDPHVDTOHLIKQLRQHFAML 537
Db 440 PA-----QKHQLOQLSOFVLM 456

RESULT 12
CC27_YEAST STANDARD; PRT; 758 AA.
ID CC27_YEAST
AC P38042;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division control protein 27.
DE CDC27 OR SNB1 OR YBL084C OR YBL0718.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Saccharomyces cerevisiae.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=5288C;
RX MEDLINE=96076635; PubMed=7502586;
RA Obermaier B., Gassenhuber J., Piravandi E., Domdey H.;
RT "Sequence analysis of a 78.6 kb segment of the left end of
RT Saccharomyces cerevisiae chromosome II.";
RL Yeast 11:1103-1112(1995).
RN [2]
RP PARTIAL SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=92306611; PubMed=1819514;
RA Slikowski R.S., Michaud W.A., Wootton J.C., Boguski M.S., Connelly C.,
RA Hietar P.A.;
RT "TPR proteins as essential components of the yeast cell cycle.";
RL Cold Spring Harb. Symp. Quant. Biol. 56:663-673(1991).
RN [3]

```

RP SUBUNITS.
RX MEDLINE=95009933; PubMed-7925276;
RA Lamb J.R., Michaud W.A., Sikorski R.S., Hieter P.A.;
RT "cdc16p, Cdc23p and Cdc27p form a complex essential for mitosis.";
RL EMBL J. 13:4321-4328(1994).
CC -1- FUNCTION: EXECUTES ESSENTIAL MITOTIC FUNCTIONS NEAR THE
CC METAPHASE/ANAPHASE TRANSITION.
CC SUBUNIT: CDC16, CDC23 AND CDC27 FORMS A MACROMOLECULAR COMPLEX.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE CDC27/NUC2 FAMILY.
CC -1- SIMILARITY: CONTAINS 8 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X79489; CAA56022.1; ALT_INIT.
DR EMBL: Z35845; CAA84905.1; -.
DR PIR: S45825; S45825.
DR SCD: S0000180; CDC27.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 8.
DR SMART: SM00028; TPR; 5.
KW Cell division; Cell cycle; Mitosis; Repeat; TPR repeat;
KW Nuclear protein.
FT REPEAT 154 187 TPR 1.
FT REPEAT 472 505 TPR 2.
FT REPEAT 540 573 TPR 3.
FT REPEAT 575 607 TPR 4.
FT REPEAT 608 641 TPR 5.
FT REPEAT 643 675 TPR 6.
FT REPEAT 676 709 TPR 7.
FT REPEAT 711 743 TPR 8.
FT DOMAIN 358 391 ASN-RICH.
FT MUTAGEN 613 613 G->D: IN TEMPERATURE SENSITIVE MUTANT.
SQ SEQUENCE 758 AA; 85436 MW; 8612EA4504327A02 CRC64;

Query Match 4.6%; Score 130; DB 1; Length 758;
Best local similarity 19.4%; Pred. No. 0.014;
Matches 54; Conservative 45; Mismatches 114; Indels 66; Gaps 7;
OY 162 RTPTAVTARPTTSSGSRVRLGTASMLTSPDGFILNLSRLNRYKSOGRP--LAKACL 218
DB 396 KTRPNLY-----SSGTR-----LTTSKKNPRSLIISNITL 425
OY 219 SISIFMKMLRLMIMIALSTHSOYKDM-----WKKVOIGKCYR 259
DB 426 TSDQITLP-EIMYNFALIRSSSOYNSFKAIRLFESQIPSHIKDTMPCWCLVOLKLF 484
OY 260 LGMREAEKQK-----SALKQOEMVDFPLYLAKVYVSIDQPYTALNFKQGLDRPG 312
DB 485 IINDMSLKTFNRLKQDQPARAVKDMETFLILW-----HLHDKVKSNNLNGMLDTPN 538
OY 313 EYVLLGIAIRIYEMNMNMSAAEYKKEVLKODNTHVAXIACIGSNHFSIDPEIALREYR 372
DB 539 KPTWCCIGNLISLQKHDAIKAFKATQDLPNFAVAYTLQGHHSNDSDSKAKTCYR 598
OY 373 RLDMGIYNGOLFNNLGLCCFYAQQYDMTLTSPERALS 411
DB 599 KALACDPQHNAYYGLGTSAKMLQYEFALLYFEKARSI 637

RESULT 13
TG37_MOUSE
ID TG37_MOUSE STANDARD; PRT; 824 AA.
AC Q61371;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Recessive polycystic kidney disease protein Tg737 (TgN(imorpk)737Rpw).
GN TG737 OR TGN737Rpw.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_Taxid=10090;
RX MEDLINE=94248756; PubMed-8191208;
RA Moyer J.H., Lee-Tischler M.J., Kwon H.-Y., Schrick J.J., Avner E.D.,
RA Sweeney W.E., Godfrey V.L., Cacheiro N.L., Wilkinson J.E.,
RA Moench R.P.;
RT "Candidate gene associated with a mutation causing recessive
RT polycystic kidney disease in mice."
RL Science 264:1329-1333(1994).
CC -1- DISEASE: Defects in Tg737 are the cause of recessive bilateral
CC polycystic kidney disease (PKD) with collecting duct and tubule
CC ectasia, and a liver lesion involving biliary dysplasia and/or
CC portal fibrosis.
CC -1- SIMILARITY: CONTAINS 11 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L31959; AAB59705.1; -.
DR MED: MGI:98715; TGN737Rpw.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 10.
DR SMART: SM00028; TPR; 7.
KW Repeat; TPR repeat.
FT REPEAT 196 229 TPR 1.
FT REPEAT 232 265 TPR 2.
FT REPEAT 271 304 TPR 3.
FT REPEAT 304 337 TPR 4.
FT REPEAT 337 370 TPR 5.
FT REPEAT 370 403 TPR 6.
FT REPEAT 403 436 TPR 7.
FT REPEAT 436 469 TPR 8.
FT REPEAT 469 502 TPR 9.
FT REPEAT 502 535 TPR 10.
FT REPEAT 535 568 TPR 11.
SQ SEQUENCE 824 AA; 92983 MW; BC214BA4B8FE3B5D CRC64;

Query Match 4.6%; Score 130; DB 1; Length 824;
Best local similarity 19.9%; Pred. No. 0.015;
Matches 86; Conservative 62; Mismatches 152; Indels 132; Gaps 18;
OY 122 NOTGPGSAVR-----PIQAGRI-TGF--LRPSQSRP--GTMEQAIR 162
DB 33 NDTGF-QQAVRTSHGRPPVTAKIPSTAVSRPAGYGSKTSLSMGRPMGTGTQDGA 91
OY 163 TPTAVTARPTTSSGSR--FVRLGTASMLTSP-----DGFILNLSRL--NLTYSOGRP 211
DB 92 RPTAVRAGSGSKALRSADPDLQSGRPAPLEAKNEDSPEETIRQLEKKVNLVEES 151
OY 212 KLRACISISFTMKMLRLMIMIALSTHSOYKDMWKKVOIGKCYRLGMYREAEKQK 271
DB 152 CIANSC-----GDLKLALEKAKDAGRER 175
OY 272 SALKQOEMV-----DTFLYLAHYVSIDQPYTALNFKQGLDK--FPREVTL 317
DB 176 VLVRQEQVTSPTENINDLTYSVLENLASQYSANEMVADALNTVOYIVKKNFSGNARLK 235
OY 318 CGIARIYEMNMNMSAAEYKKEVLKODNTHVAXIACIGSNHFSIDPEIALREYRRL 377

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL: I23076; AAA64794.1; -
DR EMBL: 249701; CAA89730.1; -
DR SGD: S0002652; PEK5.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 4.
DR SMART: SM00028; TPR; 4.
KW Peroxisome; Repeat; TPR repeat; Transport; Protein transport.
FT REPEAT 64 97 TPR 1.
FT REPEAT 313 346 TPR 2.
FT REPEAT 347 380 TPR 3.
FT REPEAT 381 418 TPR 4.
FT REPEAT 419 456 TPR 5.
FT REPEAT 457 490 TPR 6.
FT REPEAT 491 524 TPR 7.
FT REPEAT 525 558 TPR 8.
SQ SEQUENCE 612 AA; 69324 MW; 553251971E0BFD8D CRC64;

Query Match 4.5%; Score 127; DB 1; Length 612;

Best Local Similarity 22.6%; Pred. No. 0.017; Mismatches 107; Indels 72; Gaps 10;

Matches 66; Conservative 47; Mismatches 107; Indels 72; Gaps 10;
QY 265 EAERKQKSAALKQ-EMVDPEFLAKYVVSIDQPVTLNLKQGLDKPFGVTLGIGIARI 323
DB 331 EAALAEFAAVKRPDHDVAVLRGLGVOTONEKELNGISALEECIKLDPKMLKAKTLAIS 390
QY 324 Y-EEAMNMSAA-----EYKEVLKQDN-----THVXALACIG-----SN 357
DB 391 YINEGYDMSAFYMLDKWAETKYPEIWSRIKQDDKFKQEKGFTHIDMNAHITKQFIOLAN 450
QY 358 HFSYDQPEILAFYRRLQWGIYNGOLFNNLGLCCFYAQOQYDMTLTSFERALSLAENEEE 417
DB 451 NLSTIDPEIOL-----CLGLLFTYTKDDPKTIDCFESALKRYNPDEL 492
QY 418 AADVWYNLGHVAVGIDPTNLAHQCFLALVNNNNHAEVNNLAVLEMRKGHVEQARALLO 477
DB 493 ---MMNRLGASLANSNSSEALQAYHRALQKPSFVARARYNLAVSSMNICGFKKAAGYLL 549
QY 478 TASSLAPHWEPHFNFATISDKIGD-----LQRSYVAQKSE 514
DB 550 SVLS---MHEVNTN---NKKGVGSLNTYNDVYIEFLKRYFIAMRND 592

Search completed: July 17, 2002, 16:46:09
Job time: 150 sec

THIS PAGE BLANK (USPTO)